Situated Displays in Telecommunication

Ye Pan

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Computer Science

University College London

September 2, 2015

I, Ye Pan, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

In face to face conversation, numerous cues of attention, eye contact, and gaze direction provide important channels of information. These channels create cues that include turn taking, establish a sense of engagement, and indicate the focus of conversation. However, some subtleties of gaze can be lost in common videoconferencing systems, because the single perspective view of the camera doesn't preserve the spatial characteristics of the face to face situation. In particular, in group conferencing, the 'Mona Lisa effect' makes all observers feel that they are looked at when the remote participant looks at the camera.

In this thesis, we present designs and evaluations of four novel situated teleconferencing systems, which aim to improve the teleconferencing experience. Firstly, we demonstrate the effectiveness of a spherical video telepresence system in that it allows a single observer at multiple viewpoints to accurately judge where the remote user is placing their gaze. Secondly, we demonstrate the gaze-preserving capability of a cylindrical video telepresence system, but for multiple observers at multiple viewpoints. Thirdly, we demonstrated the further improvement of a random hole autostereoscopic multiview telepresence system in conveying gaze by adding stereoscopic cues. Lastly, we investigate the influence of display type and viewing angle on how people place their trust during avatar-mediated interaction. The results show the spherical avatar telepresence system has the ability to be viewed qualitatively similarly from all angles and demonstrate how trust can be altered depending on how one views the avatar. Together these demonstrations motivate the further study of novel display configurations and suggest parameters for the design of future teleconferencing systems.

Acknowledgements

My research is supported by the FP7 EU collaborative project BEAMING (248620), the UCL EngD VEIV center and the China Scholarship Council.

First and foremost, I would like to thank my first advisor, Prof. Anthony Steed. He provided me with an excellent opportunity to study in his lab and enjoy life in London. When I started my PhD, I barely had any experience in computer science, as my first degree is in electronic engineering and Japanese. Prof. Anthony Steed taught me from the beginning, introducing me to the virtual environment display technology. For the last four years, he guided me in many ways, from imparting his erudite knowledge, supporting me to attend conferences, to developing my research and programming skills. Without his step-by-step instructions and encouragements, this PhD would not have been achievable. Also, I would like to thank my second advisor, Dr. Tim Weyrich for his great suggestions in my first year viva and Beaming Project meetings.

Dozens of people have helped and taught me immensely at the VECG group. I wish to thank Prof. Mel Slater, Prof. Jan Kautz, Dr Simon Julier, Dr Gabriel Brostow and Dr David Swapp who have taught me several modules in Computer Graphics, Vision and Imaging. Thanks also go to Dr William Steptoe and Dr Oyewole Oyekoya, who have provided me with hands-on working experience through all these years. Further, thanks to all these in the room 4.17, 1ES and beyond. I have been extremely lucky to have so many kind people around.

I also thank Denis Timm, Dave Twisleton and Nick Turpin in the Technical Support Group; and Richard Gamester in the Institute of Making. Thanks for helping me to build cameras and projector frameworks and teaching me about the professionalism needed when cutting and gluing materials.

Most of all, I wish to express my deepest gratitude to my mom and dad who have provided me with generous love for the last 25 years. Being university professors

themselves, my parents encouraged me to pursue my PhD. Furthermore, I wish to thank all my friends, past and present, for enriching my PhD life with much fun.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Significance of research topic

As early as 1876, the telephone was first patented by Dr. Alexander Graham Bell. Two years later, an early concept of a combined videophone and wide-screen television called a telephonoscope was conceptualized. Then, AT&T presented a demonstration of its picture phone at the World's Fair. AT&T's demonstration has significant impact on the technology and business infrastructure; however, it only had 500 users and faded away in 1974. They tried again in 1992 with the VideoPhone 2500, but that failed again as that product only lasted until 1995. Other major players who have tried in the video conferencing space include IBM, Philips, and Sony.

Today, a handful of major video conferencing players fill certain needs. A growing number of businesses have turned to video conferencing instead of face-to-face meetings to exchange documents, thoughts and ideas. This promotes enhanced efficiency, lowers overhead expenses and gives quicker results. However, in-person communication still maintains an important role in the business world. This indicates that current video conferencing designs do not adequately meet the current needs of the users.

1.2 Research problem

From psychological and linguistic studies, it is known that non-verbal behaviours, particularly, gaze direction, fulfil many functions in person to person communication [25]. For example, mutual gaze narrows the gap between humans, since "the eyes are the window to the soul." [53] Also, gaze direction is a predictor or cue of attention in multi-party communication [126].

Currently, video telecommunication systems have a limitation in presenting gaze direction, because the participant's eye direction is different from the video camera's capturing direction. The challenges in teleconferencing include:

1. Parallax effect: when the local participant looks at the image of the remote participant in the eyes, the remote participant sees an image which suggests they are being looked aside because of the displacement between the camera and the image.

2. Collapsed viewer effect (Mona Lisa effect): for group teleconferencing, when a participant looks into the camera, everyone at the local room feels that the participant looking toward them; when the participant looks away from the camera (for example, toward other participants in the meeting), no one sees the participant looking at them.

The research was guided by and addressed, the following overall motivations:

1. A variety of systems have been developed to support gaze awareness in group video conferencing, though the majority use a 2D planar display. However, those planar displays are visible from the front only.

2. Current immersive systems, such as, CAVE and head mounted display, which can replicate a life-like face to face conversation. However, real world is blocked out (i.e. user can only see the virtual world and virtual objects).

3. Some situated displays (i.e. those are small enough to situate almost anywhere in a room, but visible from all directions) which have been built. However, most of them only have a mono or stereo image which is presented on the display, thus they are currently developed for a single observer.

4. The use of autostereoscopic display technologies could support multiple users simultaneously each with their own perspective-correct view without the need for special eyewear. However, these are usually restricted to specific optimal viewing zones.

5. Gaze and trust formation on these non-planar displays have not been evaluated yet.

(a) Spherical video telepresence system (b) Spherical avatar telepresence system

(c) Cylindrical video telepresence system (d) Random hole autostereoscopic multi-view telepresence system

Figure 1.1: Sampled photos for four situated multi-view displays.

The development of modern technology, high-speed network, efficient multi-media coding standards, low-cost large plasma or LCD display and inexpensive large screen projections, provide the opportunity to investigate more natural telecommunication systems. This thesis presents designs and evaluations of a series of situated displays.

1.3 Contributions

1.3.1 Contributions to telepresence displays

We designed and built a series of situated displays which could be used in future teleconferencing. The four displays shown in Figure 1.1(a) to Figure 1.1(d). A remote user is presented in each situated display and can engage in local conversation. Local viewers are able to understand the remote user's gaze direction. These newly designed situated displays aim to achieve the following goals: low-cost, freedom from 3D glasses

(Using 3D glasses is difficult to detect gaze direction in two-way conversation), large number of observers, wide field of view and precise gaze direction in the simulated conversation. A brief introduction of this system is given in this section and further detailed explanation is presented in Chapter 3.

Figure 1.1(a) shows a spherical display to present real-time video of the remote person. We used a non-planar display, in particular a spherical display as this type of display provides the same angle of view from all directions. Because cameras are now becoming very cheap, we further used a camera array to capture the remote user, so that we can select an appropriate video of them to show. This system is developed for teleconferencing applications that only require a single observer at multiple viewpoints to see a correct perspective image of the remote person. It offers a 360° view whereas flat displays are only visible from the front.

Figure 1.1(b) shows a spherical display featuring a ray-traced view-dependent rendering method to represent the remote person as a virtual avatar. We detail a method for enabling the displayed avatar to reproduce the facial expression captured from a person in real-time and with high-fidelity. The system provides an observer with perspective-correct rendering and the nature of the display offers surrounding visibility.

Figure 1.1(c) shows a cylindrical display to present real-time video of the remote person for multiple observers. We used an array of cameras to capture a remote person, and an array of projectors to present each of them onto the cylindrical screen. The cylindrical screen reflects each image to a narrow viewing zone without crosstalk. This system allows multiple observers to see perspective-correct images of the remote person from multiple viewing directions simultaneously.

Figure 1.1(d) shows a random hole autostereoscopic display. We developed a view-dependent ray traced rendering method to represent a remote person as an avatar on the random hole display. The method allows multiple observers in arbitrary locations to perceive stereo images simultaneously. This system could be used for group teleconferencing.

Our current systems are used for asymmetric conversations, such as teaching scenarios. Systems using similar principles could be configured to support symmetric, 3-way or N-way conversations. The low cost and ease of setup make these interesting platforms for next generation video conferencing. The borderless spherical or cylindri-

cal display can be statically situated as an interesting display for virtual avatars or other content. It could also be mounted on a robot as a mobile display for telepresence.

1.3.2 Contributions to human factors

While this work's driving motivation lies in the aspiration to enhance telepresence by building novel displays, insight into the understanding of how people behave and respond when engaged in these displays is a no lesser goal. The work had evaluated the affordances of spatial interaction and interpersonal communication of such systems.

Firstly, the work empirically evaluated the effect of perspective on the user's accuracy in judging gaze direction. We found the following results:

1. We found several models and effects for predicting the distortion introduced by misalignment of capturing cameras and observer's viewing angles in video conferencing systems. Those models might be able to enable a correction for this distortion in future display configurations (Chapter 4 to Chapter 5).

2. We also found the combined presence of motion parallax and stereoscopic cues which significantly improved the effectiveness with which observers were able to assess the avatar's gaze direction. This motivates the need for stereo in future multiview displays (Chapter 6).

Secondly, the research also investigated how trust can be altered depending on how one views the remote person. Findings are as follows:

1. By monitoring advice seeking behavior, we found that if participants observe an avatar at an oblique viewing angle on a flat display, they are less able to discriminate between expert and non-expert advice than if they observe the avatar face-on (Chapter 7).

2. By preserving a virtual avatar's correct appearance and gaze direction, the spherical display is able to maintain a consistently high level of trust regardless of viewing position(Chapter 7).

Thirdly, the experiments in this research not only rely on users' self-reports, such as qualitative interviews or questionnaires, but also quantitative studies. The frameworks for evaluating those systems could be useful for the future system evaluation.

1.3.3 Contributions to graphical rendering

We developed view-dependent ray traced rendering methods to represent a remote person as an avatar on the spherical display and the random hole display, respectively. These algorithms also could be extended to other display surfaces.

1.4 Scope of thesis

This thesis is concerned with the design and evaluation of four situated multiview telepresence displays. The work investigated the use of such displays to support both object-focus and interpersonal collaboration.

As covered in Chapter 2 , there are several potential conversation scenarios which would be used in teleconferencing. However, this work is explicitly concerned with asymmetric telepresence systems.

A variety of flat multiview systems have been developed to improve several aspects of teleconferecing. Current immersive systems also can replicate a life-like face to face conversation. However, this work is focus on situated telepresence systems.

For evaluating our four novel displays, this thesis will focus on two human factors: object-focused gaze and interpersonal trust.

1.5 Publications relating to this thesis

The research that forms part of this thesis has led to several publications. Table 1.1 matches the contributions of this thesis to individual publications. Four evaluations of our telepresence displays are presented, and the chapter in which each may be found is presented in the right most column. The display and the affordance columns refer to the telepresence system used in the evaluation and its unique affordance.

1.6 Structure

Chapter 2 contextualises the research by expanding upon the motivation, the central problem addressed, and the general approach taken. This thesis looked into challenging of teleconferencing for different conversation scenarios, previously proposed solutions to this problem, and previous evaluation methods. This work is motivated by results of studies on the advantages and disadvantages of the reproduction of eye direction in teleconferencing.

Table 1.1: Publications relating to this thesis

Chapter	Display	Affordance	Publication
Chapter 4	Spherical video display (Figure 1.1(a))	Prospective-correct rendering	• **Y. Pan**, O. Oyekoya and A. Steed. A surround video capture and presentation system for preservation of eye-gaze for telepresence. *PRESENCE: Teleoperators and Virtual Environments*, MIT Press, 24-1, 2015 [79] • **Y. Pan** and A. Steed. Preserving gaze direction in teleconferencing using a camera array and a spherical display. *IEEE 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, Zurich, Switzerland, Oct 15-17, 2012 [80]
Chapter 5	Cylindrical video display (Figure 1.1(c))	Prospective-correct rendering; multiple users	• **Y. Pan**, W. Steptoe and A. Steed. Comparing flat and spherical displays in a trust scenario in avatar-mediated interaction. *ACM CHI Human Factors in Computing Systems*, Toronto, Canada, April 26-May 1, 2014 [82]
Chapter 6	Random hole autostereoscopic multiview display (Figure 1.1(d))	Prospective-correct rendering; multiple users; stereo views from arbitrary positions;	• **Y. Pan** and A. Steed. Effects of 3D Perspective on Gaze Estimation with a Multiview Autostereoscopic Display. (Under submission)
Chapter 7	Spherical avatar display (Figure 1.1(b))	Prospective-correct rendering	• **Y. Pan** and A. Steed. A gaze-preserving cylindrical multiview telepresence system. *ACM CHI Human Factors in Computing Systems*, Toronto, Canada, April 26-May 1, 2014 [81]

Chapter 3 covers four novel display systems and associated algorithms. This chapter presents the design and construction of a spherical video telepresence system, a spherical avatar telepresence system, a cylindrical video telepresence system, and a random hole autostereoscopic multiview telepresence system. These systems are capable of reproducing the gaze direction of the remote person to multiple viewers. The detailed evaluation of these systems is presented in Chapter 4 to Chapter 7.

Chapter 4 presents the evaluation of the spherical video telepresence display. We are the first to compare a situated display with a planar display in conveying gaze. We measure the ability of observers to accurately judge the target at which a user is gazing. Experiment 1, as a pilot study, demonstrated that the camera array plus sphere display can convey gaze relatively accurately. Experiment 2 compared observers' performance in different flat and spherical display conditions further, by modelling systematic biases and investigating the influence of seat and target positions.

Chapter 5 presents the evaluation of the cylindrical video telepresence system. The experiment measures the ability of multiple observers to accurately judge which target the remote person is gazing at. We compared the cylindrical video telepresence display to three alternative display configurations. The experiment demonstrates that our system can convey gaze relatively accurately, especially for observers viewing from off-center angles.

Chapter 6 presents a study on the effects of 3D perspective on gaze estimation with the random hole autostereoscopic multiview telepresence system. We evaluated this system by measuring the ability of observers with different horizontal and vertical viewing angles to accurately judge which targets the avatar is gazing at. We compared 3 perspective conditions: a conventional 2D view, a monoscopic view with motion parallax, and a stereoscopic view with motion parallax.

Chapter 7 reports on two experiments that investigate the influence of display type and viewing angle on how people place their trust during avatar-mediated interaction. The first experiment explored how interpersonal cues of expertise presented on two identical flat displays with different viewing angle affect trust. The second experiment introduced a spherical display, which has advantages over a flat display because it better supports non-verbal cues, particularly gaze direction, since it presents a clear and undistorted viewing aspect at all angles. We then compared two display types by inves-

tigating how people place their trust. Together the experiments demonstrate how trust can be altered depending on how one views the avatar.

Chapter 8 draws conclusions and gives suggestions for future work.

Chapter 2

Background

Long-distance collaboration is a fact of life for an increasing number of workers, because it reduces the need for physical travel. More relationships are being formed and maintained via teleconferecing than ever before, including supplier purchaser relationships, student-teacher relationships, or collaboration with colleagues at different locations. Current technology allows local users to communicate with remote users at almost every time and every place, capturing their expressions and delivering it in real-time to geographically separated users.

Technology designers have presented a myriad of communication tools that mitigate barriers of distance in real-time communication. However, as useful as textual and audio only technologies are, we know that our bodies do a significant amount of communication to supplement, enhance, or replace the spoken or written word. Thus, visual information is an extremely valuable communication channel. However, a single camera perspective warps some of the visual information (e.g. spatial characteristics) in current teleconferencing system. In this research, we have designed, built and evaluated four novel teleconferencing systems, based on an understanding of interpersonal communication and how people perceive images in a teleconferencing setting.

This chapter aims to contextualise the research presented in this thesis, by discussing the related work that has shaped its motivation, the problem it aims to address, and the approach it takes. The chapter is comprised of five main sections, which narrow down the focal area of research. The first section explores the reproduction of non-verbal cues in telepresence for different conversation scenarios, with a particular focus on the reproduction of gaze direction. The second section discusses different telepresence display systems, and covers the relevant literature in situated displays,

multiview displays, shape-changing displays, virtual reality systems and augmented reality systems. The third section presents related work on telepresence capture systems in both video-mediated communication and avatar-mediated communication. The fourth section explores the evaluation of telepresence displays, with a particular focus on the affordance of gaze direction and interpersonal trust. The last section summarises literature presented in the previous four sections and describes the focal area of the research.

2.1 Conversation scenarios

In face to face communication, whether it is verbal or non-verbal, conscious or unconscious, our bodies are capable of powerful expression through words that are said, a smile that is shared, or the shake of a hand. However, However, some of nonverbal cues, such as gaze directions, can be lost in the visual communication systems. The gaze distortions in video conferencing are mainly caused by two factors: parallax shift effect and collapsed viewer effect [73, 71]. The parallax shift effect occurs due to a video camera tending to be perched on top of a monitor display in a traditional video-conferencing system. This causes the user's eye direction to be different from the video camera's capturing direction. Note that the parallax shift effect can occur both horizontally and vertically. The collapsed viewer effect is where all remote participants share the same virtual viewing position of the local scene. This happens in group to group video communication systems. For example, if a participant is directly looking at the capturing camera in a remote room, all the viewers in the local room will feel that the remote participant is looking at them.

In this section, we review gaze reproduction in telepresence systems for different conversation scenarios, including two-way conversations, three-way or N-way conversation, group to group conversation, and shoulder to shoulder conversation.

2.1.1 Two-way conversation

In a two-way conversation, where only two participants at different geographical locations join the video communication, there are various methods of producing a correct gaze direction [20]. Using a half-silvered mirror [6, 1], embedding a camera in the centre of display [2], or using a transparent display could allow a video camera to capture

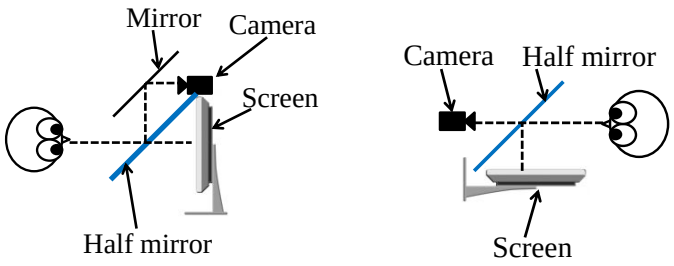
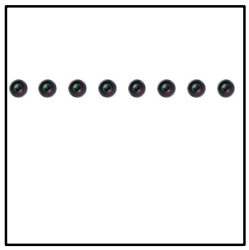


Figure 2.1: Use half silvered mirror to achieve line-of-sight .

(a) Embed one camera at the center of display



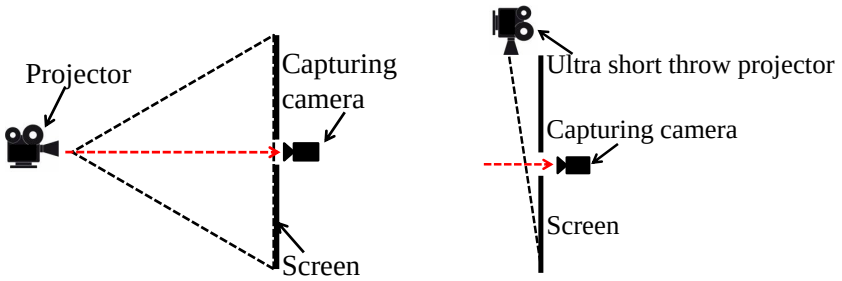
(b) Arrays of camera arranged at viewer's eye level

Figure 2.2: Embed camera in the display to achieve line-of-sight.

the participant's correct gaze direction without blocking the image on the screen.

Figure 2.1 shows two ways to place the half silvered mirrors. However, once participants are moving or not sitting in front of the display, the parallax shift effect will occur.

Figure 2.2(a) shows the design of the Mebot V4 [2], which is an example of embedding a camera in the center of display. It also has the limitation that the user cannot move during the conversation. An improved design is presented in Figure 2.2(b). There is a line of cameras which is capable of maintaining eye to eye contact even if the participants are moving horizontally. However, due to the height of users varying from individuals to individuals, the position of the camera is not always suitable for every individual. It is hard to make a hole in the center of a computer display. Comparably, making a hole in the projector screen is an accessible approach. To implement this installation, there is a problem: if place the project in front of display, the camera which is behind the screen could only detect light from the projector and cannot detect the screen in front of the camera, as shown in Figure 2.3(a). Fortunately, as shown in Fig-



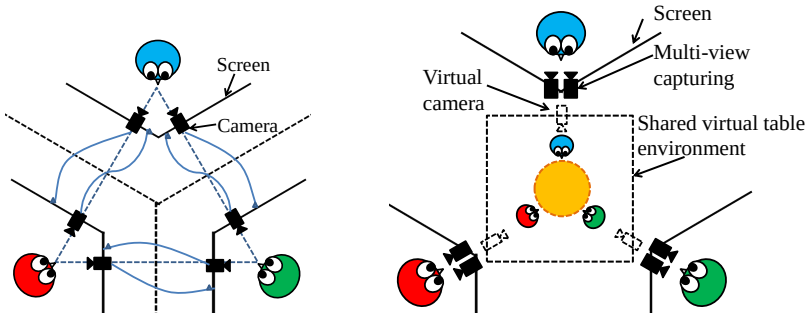
(a) Using normal projector (b) Using ultra short throw projector

Figure 2.3: Methods of arranging projector.

Figure 2.4: Transparent OLED display prototype from Samsung

ure 2.3(b), if we place a ultra short throw projector at the top of the screen, the camera is still able to detect the environment.

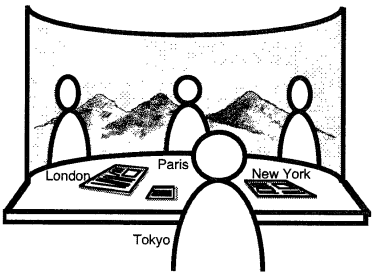
Figure 2.4 shows an example of transparent display. We could place a capturing camera behind the display. Thus, the camera could capture the correct-perspective of the user.



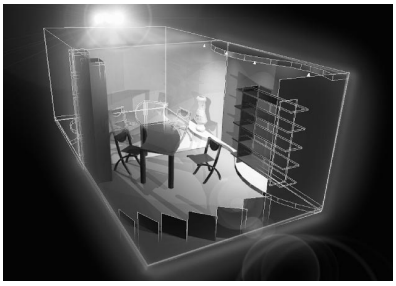
(a) Around table concept (b) Shared virtual table environment

Figure 2.5: Structure of 3-way communication network.

(a) Hydra [109]



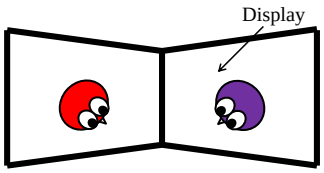
(b) MAJIC [76]



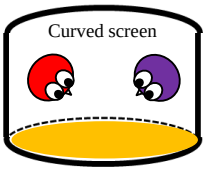
(c) TELEPORT [36]

(d) NTII [98]

Figure 2.6: Examples for 3-way teleconferencing system.



(a) Separated screen



(b) Curved screen

Figure 2.7: Three-way conversation screen.

2.1.2 Three-way or N-way conversation

For three-way or N-way conversations, more than two participants at different places link up in the conversation. Apart from considering the parallax shift effect, the structure of three-way or N-way communication network is also an essential issue. Round-table and SVTE (shared virtual table environment) are basic schemes to build a three-way or N-way communication network [105].

Figure 2.5(a) shows the round-table scheme. Many researchers have used this scheme to reproduce correct gaze direction in three-way or N-way conversations. Figure 2.6(a) shows the Hydra system [109], which placed several hydra units in front of a local user to present the videos of remote users. Figure 2.6(b) shows the MAJIC [76] system. At each site of this system, a large semi-transparent curved screen

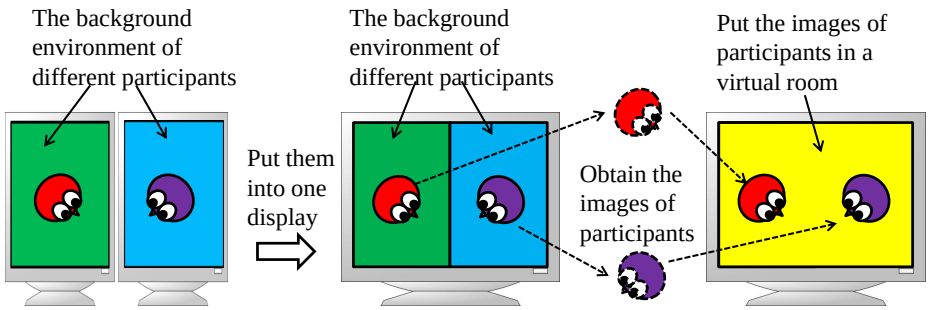


Figure 2.8: Scenario of virtual room. Left: separate screen. Middle: split screen with different background. Right: shared virtual background

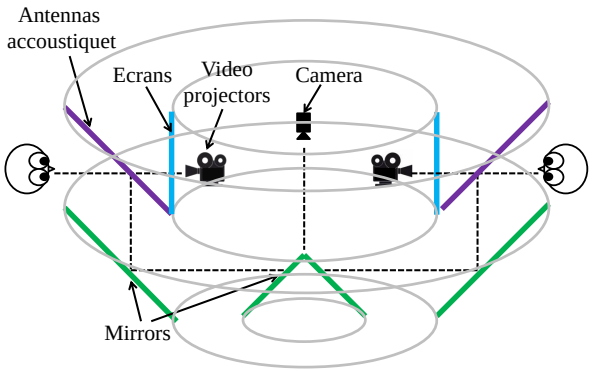
was mounted behind a normal computer terminal. In the MONJUnoCHIE system [5], a special semi-transparent display based on holographic optical elements was used.

The overall transmission bit is increased with the square of the connected sites. e.g. N $\times$ (N-1) cameras are needed for N participants [70]. Alternatively, the SVTE scheme manages to decrease the overall transmission bit by integrating generic 3D representations of the conferees into a shared virtual environment [8, 66], presented in Figure 2.5(b). In contrast to the round table scheme, such as the Hydra, this allows for the usage of efficient multicast network structures, meaning that the same generic 3D video representation is sent to all (N-1) remote destinations. The TELEPORT system in the Figure 2.6(c) [36] and NTII (National Tele-Immersion Initiative in Figure 2.6(d)) [98] utilized this idea.

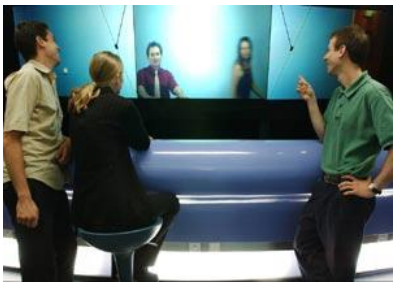
Another topic in three-way conversations is how to display all the remote users in the local site. It usually uses at least two separate windows or screens to present two remote users in the local site. There are two ways to set up those displays shown in Figure 2.7. Instead of presenting two remote participants on different screens, it is possible to segment the participants' images from their background and present these images against a virtual background, as described in Figure 2.8. This technique will lose the background information of each participant, but this is not that important in many circumstances.

2.1.3 Group to group conversation

Group to group conversation means that multiple users are collocated with an instance of the teleconferencing system. Group-to-group systems with one camera per site will necessarily distort gaze direction due to Mona-Lisa effect. When a participant looks

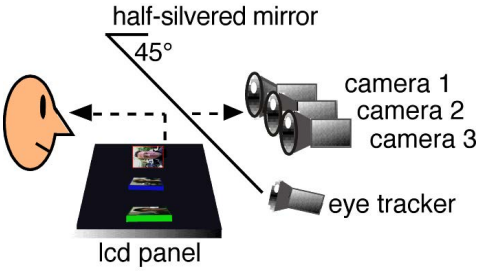


(a) schematic diagram



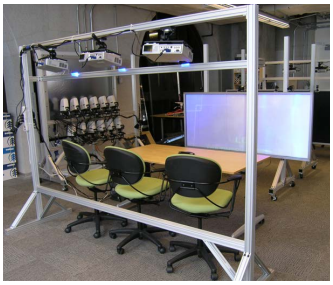
(b) front photo

Figure 2.9: Tele-presence wall. [105]

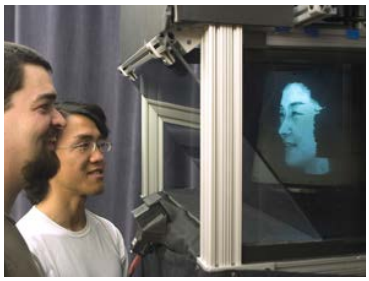


(a) Gaze-2 [127]

(b) Gcoll [112]



(c) Multiview



(d) One-to-many [46]

Figure 2.10: Examples for multi-parties teleconferencing system.

into the camera, everyone seeing their video stream sees the participant looking toward them; when the participant looks away from the camera (for example, toward other participants in the meeting), no one sees the participant looking at them.

Many of systems have been built to support correct gaze direction for group conversation. The Telepresence Wall [23] in Figure 2.9 is an example of a display used to support two groups at two sites. Figure 2.10(a) shows the GAZE-2 [127] that uses an eye-controlled camera direction to ensure parallax free transmission of eye contact. Gcoll [112] in Figure 2.10(b) supported mutual gaze as well as partial gaze awareness

for all participants with modest technical requirements, e.g. notebooks with two USB cameras for each user. These systems only work correctly and provide their affordances when used with one participant per site. This will be a problem with any system based on viewer-independent displays. In real physical space, different users do not share the same view of others. Recent systems provides a practical solution to this problem, using a custom view-dependent display. Figure 2.10(d) shows a one-to-many 3D video teleconferencing system [46]. The remote user's face is scanned in 3D at 30Hz and transmitted in real time to an auto-stereoscopic horizontal parallax 3D display, displaying it over more than 180° field of view observable to multiple views. MultiView [71] in Figure 2.10(c) accomplishes reproduction of eye gaze in group to group conversation by capturing unique and correct perspectives for each participant. It uses one of many cameras and simultaneously projecing each of them onto a directional screen that controls who sees which image.

Building on previous research, this thesis introduced several view-dependent displays to support correct gaze direction for group conversation.

2.1.4 Shoulder to shoulder conversation

Shoulder to shoulder conversations give more attention to the users' environment. It is particularly focused on representing a remote participant as a visitor to join local conversation.

As discussed above, many telepresence systems have been built to improve different videoconferencing scenarios, though the majority use planar displays. However, those planar displays are only visible from the front. The scope of this thesis is to focus on the one-way teleconferencing scenario, as an evaluation of our displays. Nevertheless, previous researches using flat displays for two-way conversation, three-way or N-way conversation, group to group conversation and shoulder to shoulder conversation scenarios are important for the future development of displays.

2.2 Display systems

In this following section, we first outline state of art situated displays and multiview displays, which shapes the motivations of this thesis. We then discuss related displays and their features that could be used in teleconferencing.

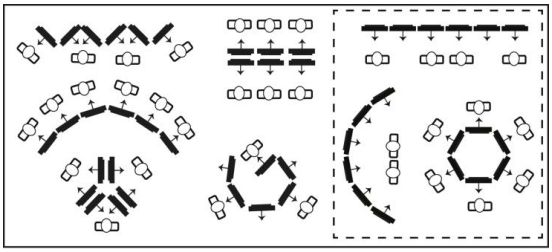
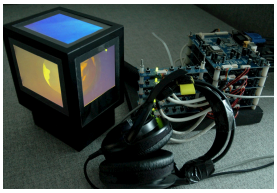


Figure 2.11: An overview of possible 6-chained display configurations. [122]

2.2.1 Situated display



(a) pCubee [114]

(b) Sphere-Avatar [78]

(c) BiReality [47]

(d) Tele-Human [50]

Figure 2.12: Examples for situated teleconferencing system.

There are different kinds of non-flat display surfaces [122], particularly, situated displays, such as spherical displays and tubular displays. These situated displays are small enough to situate almost anywhere in a room, and visible from larger range of directions than flat displays.

The BiReality system [47] uses a teleoperated robotic surrogate to provide an immersive telepresence system for face-to-face interactions. It consisted of a display cube at a user's location and a surrogate in a remote location. Both the remote participant and the user appeared life size to each other. The display cube provided a complete 360° surround view of the remote location and the surrogate's head displayed a live video of the users head from four sides. By providing a 360° surround environment for both locations, the user could perform all rotations locally by rotating his or her body. Horizontal gaze is best preserved for the user as seen by remote participants when the user is looking into the cameras in the corner of the display cube, and is sloppier when the user is looking at the center of a screen.

SphereAvatar [78] represents a remote user as an avatar on a spherical display which is able to accurately convey head gaze. In order to correct gaze distortion, flat displays either use a half mirror which will reduce the video quality and increase the display complexity, or embed the camera in the centre of the display which will block the display image. Spherical displays project the image from the bottom of the display. In this thesis, our spherical video telepresence system (see Section 3.1) extends the work of SphereAvatar [78]. We use a surround camera array to reproduce the real time video of the remote participant instead of an avatar in order to improve reproduction fidelity and preserve eye gaze. Additionally, different from the SphereAvatar which used the polygonal rendering approach to represent remote person, our spherical avatar telepresence system (see Section 3.2) used a ray tracing engine which could provide higher quality images with less distortion.

TeleHuman [50] provides 360° motion parallax with stereoscopic life-sized 3D images of users, using a lightweight approach. Motion parallax is provided via perspective correction that adjusts views as users move around the display. Stereoscopy is provided through shutter glasses worn by the user. The system uses ten Microsoft Kinects for capturing 3D video models of the user in 360°. Telehuman is a reconstruction system, whereas we focus on spatial video transmission.

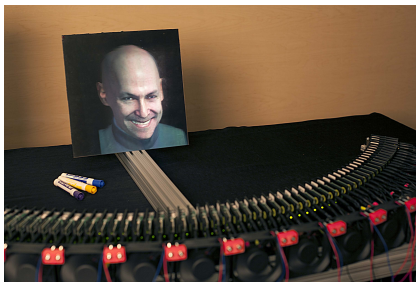
2.2.2 Autostereoscopic display

Depth perception, or 3D perception, can add a lot to the feeling of immersiveness in many applications, such as, 3D teleconferencing. For conventional stereo display, special glasses, such as colour glasses, polarizer glasses and shutter glasses, are widely used for stereoscopic 3D displays. These glasses-based technologies are not dependent on the viewing angle and they are extremely flexible. However, these displays would require the use of 3D glasses, which is cumbersome and difficult to support eye contact perception in two way teleconferencing.

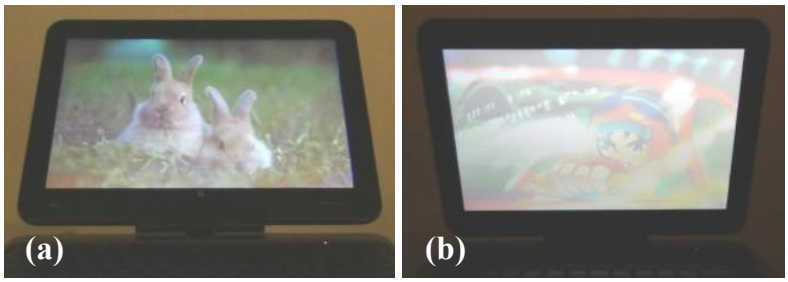
Autostereoscopic displays, presenting a 3D image to a viewer without the need for glasses or other encumbering viewing aids, can be used to improve the teleconferencing experience. These display types include holographic, volumetric, or parallax barrier.

Holographic displays [62] output a partial light-field, computing many different views simultaneously. This type of display has the potential to allow many observers

(a) Seelinder [135]



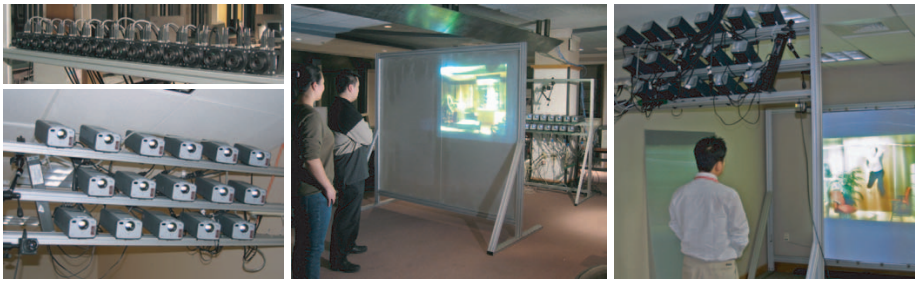
(b) 3D facial display [68]



(c) TNLCD

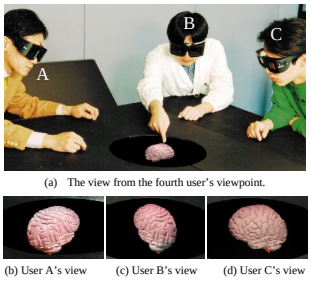
(d) Varrier [101]

(e) Dynallax [86]

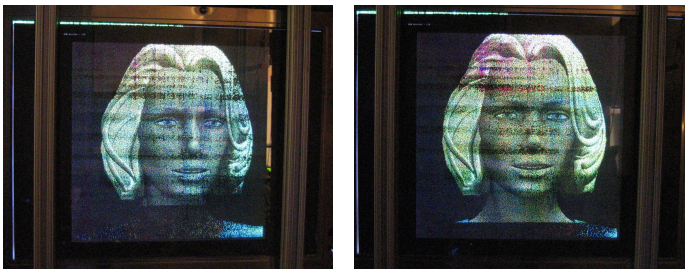


(f) 3D-TV [64]. Left: Array of 16 cameras and projectors. Middle: Rear-projection 3D display with double-lenticular screen. Right: Front-projection 3D display with single-lenticular screen.

Figure 2.13: Examples for multi-view teleconferencing system.

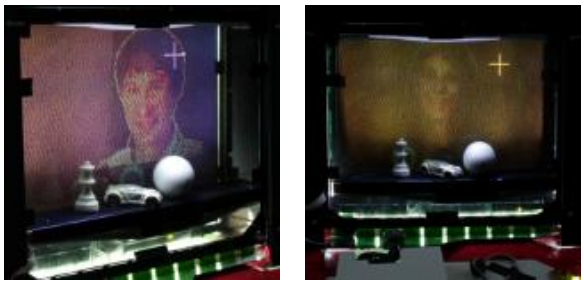


(a) IllusionHole [52]



(b) RHD [69]

(c) TAD [134]



(d) MUSTARD [48]

Figure 2.14: Examples for random hole display

to see the same object simultaneously, but of course it requires far greater computation than is required by a two-view stereo for a single observer. Generally only a 3D lightfield is generated, reproducing only horizontal, not vertical parallax.

Traditional volumetric displays do not create a true lightfield, since volume elements do not block each other [17]. The effect is of a volumetric collection of glowing points of light, visible from any point of view as a glowing ghostlike image.

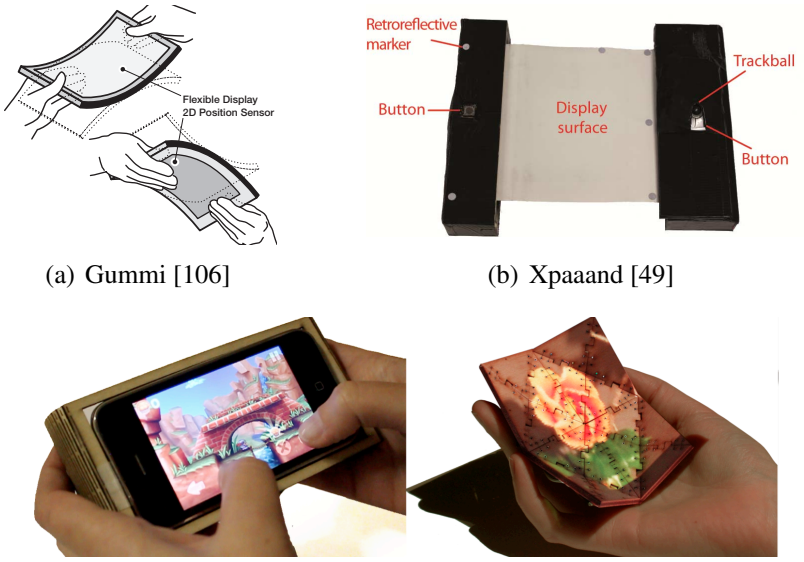
Parallax-based displays based on barriers or lenticular lens sheets provide a relatively simple and inexpensive solution for autostereoscopy. Parallax barrier displays occlude certain parts of the screen from one eye while allowing the other eye to see them. A lenticular screen is a sheet of cylindrical lenses while a parallax barrier is a flat film composed of transparent and opaque regions.

Parallax barrier displays include Perlin et al.'s autostereoscopic display [85], Varrier [101], and Dynallax [86]. In particular, Seelinder(see Figure 2.13(a)) [135] is a 3D video display technique that allows multiple viewers to see 3D images from a 360° horizontal arc without wearing 3D glasses. This technique uses a cylindrical parallax barrier and a one-dimensional light source array. This gives us an inspiration to design and evaluate our cylindrical video telepresence display (see Section 3.3).

Lenticular displays include the MERL display [64] and Kooima et al.'s work [56]. Additionally, Kim et al. proposed another approach enabling concurrent dual views on twisted-nematic LCD screens, by exploiting a technical limitation of these LCD screen [51]. In particular, 3D TV(see Figure 2.13(f)) [64] presents a system for real-time acquisition, transmission, and high-resolution 3D display of dynamic multiview TV content. This system consists of an array of cameras, clusters of network-connected PCs, and a multi-projector 3D display. Multiple video streams are individually encoded and sent over a broadband network to the display. The 3D display shows high-resolution stereoscopic color images for multiple viewpoints without special glasses. In our spherical video telepresence system (see Section 3.1) and cylindrical video telepresence system (see Section 3.3), we used a similar video capture and display network, but the displays are different.

However, neither autostereoscopic displays nor conventional stereo displays support both vertical motion parallax and multiple arbitrary views. Firstly, most conventional AS displays do not offer multiuser motion parallax (multiple distinct views) along the vertical direction. Integral imaging displays using a 2D array of lenslets could generate fullparallax autostereo images, but these have a limited viewing angle and low resolution. Therefore, it would be difficult to provide correct-perspective views for observers with different heights. With regular multi-user autostereoscopic displays, untracked viewers must remain in certain viewing areas or they will see incorrect imagery or the same imagery as other viewers. In autostereoscopic display systems with user tracking, multiple viewers are usually not supported because individual display pixels will be seen from multiple views. These can be difficult to use in group teleconferencing.

Recently, an interesting approach to build multi-view displays is based on viewing the data through a hole-mask that is placed at a certain distance from the data to serve as a barrier that mediates the view for different users. Kitamura et al.'s Illusion Hole uses a display mask which has a hole in its center. [52]. Naschel et al.'s random hole display prototype extends their approach by using a randomized hole distribution parallax barrier [69]. The random hole display design eliminates the repeating zones found in regular barrier and lenticular autostereoscopic displays, enabling multiple simultaneous viewers in arbitrary locations [69]. Gu et al. demonstrate a full multi-user multi-view



(a) Gummi [106] (b) Xpaaand [49]

(c) Morphees [97]

Figure 2.15: Examples for shape-changing displays.

system using this concept with their Tabletop Autostereoscopic Display [134]. Instead of using a static hole-mask, Karnik et al.'s MUSTARD uses a dynamic random hole mask allowing coverage of the entire screen by constantly changing the hole-mask from frame to frame [48].

Whilst autostereoscopic and multiview capabilities of a random hole display are novel, the effectiveness of using the random hole display for telepresence is not yet clear. We run an experiment to demonstrate that the random hole display can convey gaze relatively accurately, particularly for group conferencing (see Chapter 6).

2.2.3 Shape-changing display

Shape changing displays are also an interesting type of non-planar display. Recent displays have been built with the aim of using some shape-changing interface qualities to enhance our interaction with digital information.

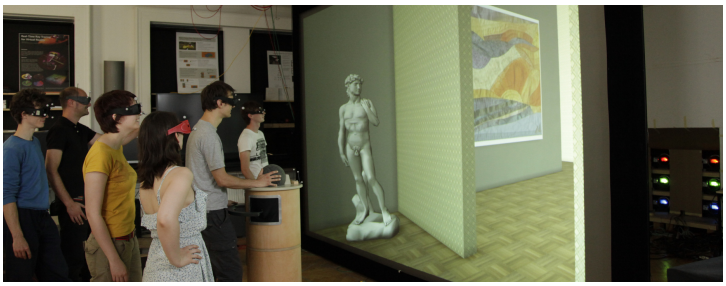
Figure 2.15(a) shows Gummi [106], which introduced a set of interaction techniques for bendable displays, which support scrolling and zooming. Evaluations of the prototype demonstrated Gummi interaction techniques to be feasible, effective and enjoyable.

Xpaaand [49] in the Figure 2.15(b) provides for dynamically resizing the mobile device and its display. The evaluation of this display showed that physical resizing

(a) VIRTUE [104]

(b) Blue-C [40]

(c) Office of the future [90]



(d) C1x6 [57]

Figure 2.16: Examples for virtual environment teleconferencing system.

of the screen real state creates a rich physical experience and can effectively improve interaction with handheld devices.

Figure 2.15(c) shows Morphees [97], that are self-actuated flexible mobile devices adapting their shapes on their own to the context of use in order to offer better affordances.

2.2.4 Virtual reality systems

The art of immersive displays can be traced back to Backer's panorama, which presented a wide vista onto a completely circular surface in correct perspective. It was so convincing that was able to trick the spectators to believe this reproduced real world is genuine [105]. Then, following the Cinerama, numerous film formats such as IMAX, 3D IMAX, and Omnimax bring distant, exciting worlds within the partici-

pants' grasp [58].

Virtual environment systems, including the HMD, BOOM and CAVE provide the users with a strong sense of presence, by their multisensory stimulation, immersive characteristics and real-time interactivity [28]. Systems such as VIRTUE (virtual team user environment) [104], im.point (immersive meeting point) [119], Blue-C [40] and Office of the Future [90] are effective ways to simulate face to face conversations by applying the concept of a shared environment [41]. TELEPORT system [36] and NTII (National Tele-Immersion Initiative) [98] utilized the SVTE concept described above to preserve gaze direction in three-way or N-way conversations. The commercial available Oculus Rift HMDs gives the user the impression of being inside of a complete virtual world.

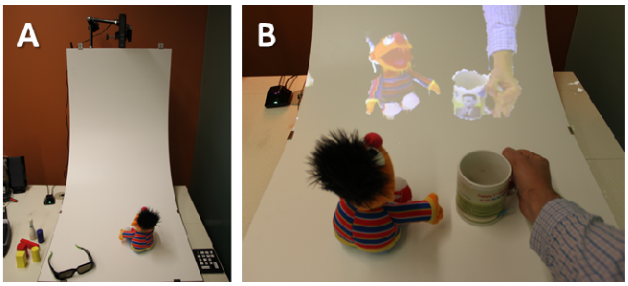
However, these systems need sophisticated equipment, such as complex display mountings, special tracking devices, etc.

2.2.5 Augmented reality systems

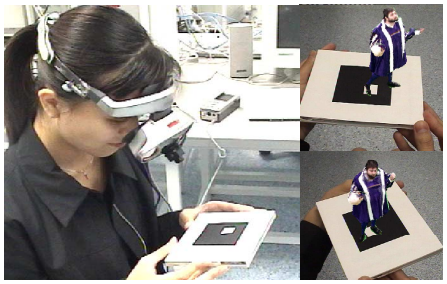
Augmented reality (AR) techniques can be used to develop fundamentally different interfaces for face-to-face and remote collaboration because AR provides seamless interaction between real and virtual environments; the ability to enhance reality; the presence of spatial cues for face-to-face and remote collaboration; support of a tangible interface metaphor; and the ability to transition smoothly between reality and virtuality. A variety of augmented reality systems have been built to develop effective face-to face collaborative computing environments [15, 9].

Figure 2.17(a) shows the MirageTable [14] which is instrumented with a single depth camera, a stereoscopic projector, and a curved screen. The authors illustrate these unique capabilities through three application examples: virtual 3D model creation, interactive gaming with real and virtual objects, and a 3D teleconferencing experience (This not only presents a 3D view of a remote person, but also a seamless 3D shared task space). They also evaluated the user's perception of projected 3D objects in their system, which confirmed that users can correctly perceive objects even when users are projected over different background colours and geometries.

Figure 2.17(b) shows a real-time 3-D augmented reality videoconferencing system [87]. With this technology, an observer sees the real world from his viewpoint, but

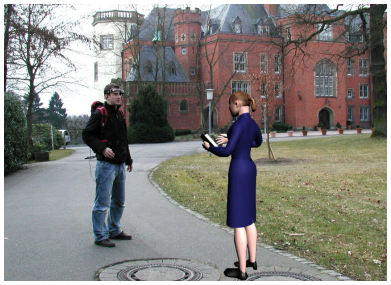


(a) MirageTable [14]



(b) 3D Live [87]

(c) AR-Rift



(d) MARA [102]

Figure 2.17: Examples for augmented reality teleconferencing system.

modified so that the image of a remote collaborator is rendered into the scene. When this view is superimposed upon the real world, it gives the strong impression that the collaborator is a real part of the scene.

Figure 2.17(c) shows the AR-Rift, a low-cost video see-through AR system using an Oculus Rift and consumer webcams. This system could also be used in the teleconferencing.

Figure 2.17(d) shows the MARA [102], which is a mobile augmented reality-based virtual assistant.This system presents a first step to integrate an anthropomorphic assistant with an AR information and navigation system.

(a) Mebot V4 [2]

(b) PRoP [84]

(c) Anybots' QB and VGo [125]

(d) Geminiod HI-1 [100]

(e) Animatronic Shader Lamps Avatars (SLA) [61]

Figure 2.18: Examples for tele-presence robot.

2.2.6 Telepresence robots

Mobile telepresence robots, such as, MeBot V4 [2], PRoP [84], Anybots'QB and the VGo [125], allow a remote user to control the robot's movement around a space while the user converses with other users in that space. These devices tend to have a built-in flat screen to display a video stream of the remote user. Using these telepresence robots, remote co-workers can wander the hallways and engage in impromptu interactions, increasing opportunities for connection in the workplace [59]. Since mobility is the characteristic that differentiates mobile telepresence robots from video conferencing technologies, we could potentially integrate a spherical display into a robotic platform.

Humanoid robotics focus more on better conveyance of a person's remote physical

presence. Geminiod HI-1 [100] was developed to closely resemble a specific human. Animatronic Shader Lamps Avatars (SLA) [61] use the technique where an image of an object is projected onto a screen whose shape physically matches the object. It uses cameras and projectors to capture and map the dynamic motion and appearance of a real person onto a humanoid animatronic model. Those humanoid robots can potentially be used to represent specific visitors at a destination but they are limited in terms of their flexibility in representing other users.

2.3 Capturing systems

The previous section discussed the displays in telepresence systems. This section contextualises the technical aspects of the corresponding capture systems.

Video conferencing is the most established and accessible forms of audio-visual remote interaction for dyadic and small group communication. However, even minor physical movement of a user may introduce parallax between camera position and video display resulting in loss of gaze awareness [95]. The 2D nature of a standard video constrains the rich spatial cues common to collocated interaction such as depth, resolution, and field of view [116]. In regard to spatiality, videoconferencing has proven to be more similar to audio conferencing than to unmediated interaction [131]. In order to capture perspective-correct videos, one can record the remote person by camera arrays [120].

Recently, avatar-mediated communication, where a remote person is represented by a graphical humanoid, has increased in prevalence and popularity as an emerging form of visual remote interaction [32]. The avatar represents the presence and activities of a remote user and can be visualized using standard displays or projection surfaces in the local room with perspective-correct graphical rendering via head tracking of the local user [94]. Avatars are capable of eliciting appropriate responses from observers (see e.g. [11], [116]).

The following sections detail capturing methods for video-mediated communication and avatar-mediated communication, which are the two mediums of visual telecommunication with which the work in this thesis is concerned.

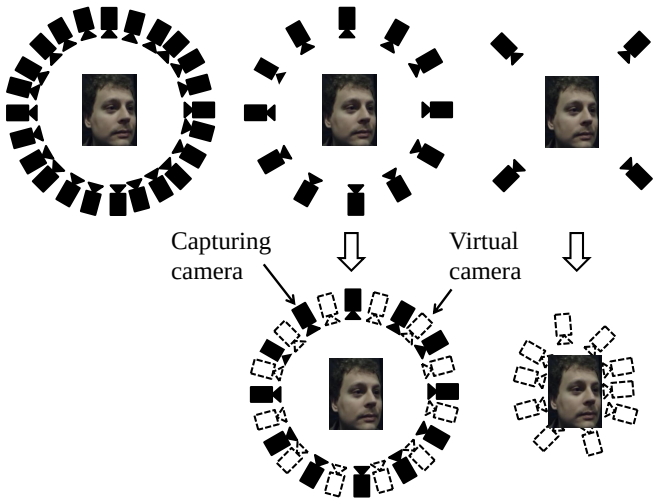


Figure 2.19: Free view point image generated with different camera densities. Left: Shows the case of very dense or continuous camera configuration, such as very dense ray space. In this case, any viewpoint image can be easily be obtain by collecting the rays that pass the viewpoint from difference camera. Middle: Shows the case of dense camera configuration. In this case, undetected rays are generated by interpolation. Right: Shows the case where camera configuration is so sparse that the interpolation of undetected rays is difficult. This is model based case. A 3D model of the object is made and texture is mapped on the surface of the object.

Table 2.1: Summary of video manipulation approaches

Method	**Explanation**	
Show video	Simply presents the video stream.	
Free view point	A scene is captured by a set of cameras	The camera density is very high
		The camera density is moderately high
		The camera density is low
Reconstruction of 3D model	The 3D context of each user's physical environment is lost.	
Segmented video	Real-time separation of foreground from background [35].	

2.3.1 Video

The arrangement of video cameras can be divided into three basic categories: fixed camera, moving camera and camera arrays. The diverse video handling techniques with appropriate examples are summarized in Table 2.1. The most straightforward one is fixed camera (e.g. the original line-of sight system in Figure 2.1). This kind of system is simple and inexpensive. However, since the camera cannot move, it limits the user to a specific position. For a moving camera, the camera's position changes accord-

ing to the direction given by the user's eye. One of the representative examples of a moving camera system is a tele-presence robot. The last type of arrangement is using camera arrays. Many cameras $(10-100+)$ that cannot move; the proper cameras are picked and edited based on the position of the user's head or eyes (see Figure 2.19). For very dense camera configuration, view generation is simply by selecting a camera image or by collecting pixels from camera image. These systems include NHK system [7], 1D integral image 3D display system [43]. For dense camera configuration, view generation needs some processing. These system include FTV(Free viewpoint TV) [121], Birds eye view system [108], Light field camera system [130], Surface light field camera system [24], EyeVision, 3D-TV [64], Free viewpoint play. For sparse camera configuration, intermediate view can be generated by detecting model in the scene. These system include 3D room [99], 3D Video [63], Multi-texturing [113].

2.3.2 Avatar

Conversation includes spoken language, including words and contextually appropriate intonation marking topic and focus; facial movements, including lip shapes, emotions, gaze direction, head motion; and hand gestures, including hand shapes, points, beats, and motions representing the topic of accompanying speech. Video mediated communication can provide a rich mode of visual interaction, in which the users can see and hear each other in real time and communicate using both verbal and non-verbal cues such as speech, gaze and facial expression. In avatar-mediated interaction, it is important to capture high fidelity avatars. Without all of these verbal and nonverbal behaviours, one cannot have realistic, or at least believable, avatars.

Facial animation approaches could be grouped into two groups, those based on geometry manipulation and those based on image manipulation [88]. Geometry manipulation refers to manipulation of 3D models that consist of vertices in space forming polygons and thus representing the surface of the model. The geometry manipulation methods include key-framing, parameterization, pseudo-muscle methods, and physics-based methods. Image manipulation refers to 2D images or photos that are morphed from one to another in order to achieve a desired animation effect. The image manipulation methods include morphing and blendshapes.

In this research, we are interested in the performance driven animation (also re-

ferred to as expression mapping) which assumes a performer and makes appropriate use of both geometry-based and image-based techniques to do the animation. However, most methods typically require complex acquisition systems and substantial manual post-processing. As a result, creating high-quality character animation entails long turn-around times and substantial production costs. A full review of these performance driven animation systems is beyond the scope of this thesis and we refer to [132] for a more detailed discussion.

In particular, with recent developments in gaming technology, such as the Nintendo Wii and the Kinect system of Microsoft, Faceshift® demonstrated a high-fidelity and real-time parametric reconstruction of facial expression method without the use of face markers, intrusive lighting, or complex scanning hardware [129]. The user is recorded in a natural environment using a non-intrusive, commercially available 3D sensor. The simplicity of this acquisition device comes at the cost of high noise levels in the acquired data. To effectively map low-quality 2D images and 3D depth maps to realistic facial expressions, they introduced a novel face tracking algorithm that combines geometry and texture registration with pre-recorded animation priors in a single optimization. Formulated as a maximum a posteriori estimation in a reduced parameter space, their method implicitly exploits temporal coherence to stabilize the tracking. We used this capturing method in our spherical avatar telepresence system (see Section 3.2) and random hole autostereoscopic multiview telepresence system (see Section 3.4), as this method only requires a single depth camera.

More recently Li et al. [60] introduced a real-time markerless facial animation framework. This method can be instantly used by any subject, without training (comparing to [129]), and ensures accurate tracking using an adaptive PCA model based on correctives that adjusts to the users expressions on-the-fly. We plan to use this technology in future research.

2.4 Evaluation methods

In an extensive review of studies of distributed and collocated work, Olson and Olson [77] identified relevant factors that make a difference in these work contexts (see Table 2.2). To further understand collaboration systems, three aspects need to be taken into consideration: person space, task space and reference space. Person space is usu-

Table 2.2: How well today's and future technologies can support the key characteristics of collocated synchronous interactions.

Characteristic	Description	Today	Future
Rapid feedback	As interactions flow, feedback is as rapid as it can be. Quick corrections possible when there are noticeable misunderstandings or disagreements.		well supported
Multiple channels	Information among participants flows in many channels, including voice, facial expressions, gesture, body posture, and so on. There are many ways to convey a subtle or complex message; also provides redundancy.	poorly supported	well supported
Personal information	The identity of contributors to conversation is usually known. The characteristics of the source can be taken into account.	poorly supported	well supported
Nuanced information	The kind of information that flows is often analog or continuous, with many subtle dimensions (e.g., gestures). Very small differences in meaning can be conveyed; information can easily be modulated.	poorly supported	well supported
Shared local context	Participants have a similar situation (time of day, local events). A shared frame on the activities; allows for easy socializing as well as mutual understanding about what is on each others minds.		
Informal "hall" time	Before and after Impromptu interactions take place among subsets of participants on arrival and departure. Opportunistic information exchanges take place, and important social bonding occurs.	poorly supported	poorly supported
Coreference	Ease of establishing joint reference to objects. Gaze and gesture can easily identify the referent of deictic terms.		poorly supported
Individual control	Each participant can freely choose what to attend to and change the focus of attention easily rich, flexible monitoring of how all of the participants are reacting to whatever is going on		poorly supported
Implicit cues	A variety of cues as to what is going on is available in the periphery. Natural operations of human attention provide access to important contextual information.		poorly supported
Spatiality of reference	People and work objects are located in space Both people and ideas can be referred to spatially; "air boards".		poorly supported

ally achieved with video and audio connections. The evaluations often consider how well those systems could simulate users' verbal and non-verbal cues to present expression, trust. The task space is where the work appears typically realized through a shared workspace application. The reference space is where remote parties can use body language to refer to the work, including gaze direction, pointing, gesturing etc. This is often realized as mouse pointers, though also as video embodiments of arms [118].

In accordance with the aim of this research, we review related work in the affordances of gaze direction and interpersonal trust of telepresence systems. We then describe the development of the evaluation framework in detail and statistical analysis method that are used in this thesis.

2.4.1 Gaze

Gaze, attention, and eye contact are important aspects of face to face conversation. They help create social cues for turn taking, establish a sense of engagement, and indicate the focus and meaning of conversation [25]. However, perceiving gaze direction is difficult in most teleconferencing systems and hence limits their effectiveness [71]. In this section, we first look into several human factors studies in the perception of head and eye gaze direction, which inform the design and evaluation of telepresence systems. We then discuss previous evaluation frameworks used in evaluating telepresence systems. The last part of this section motivates the research problem by discussing the affordances of previous telepresence systems.

2.4.1.1 Perception of head and eye gaze direction

Early work indicates that gaze direction may be perceived by both the direction in which the head is oriented and the eyes' positions relative to the head [37]. Other research has focused on studies in which the eyes and the head were counter-rotated to varying degrees while maintaining fixation on the subject [37, 4]. These studies consistently showed an interaction between eye and head position in the perception of gaze direction. Gibson et al. [37] examined three head gaze conditions: head to front, left and right. In each condition, an observer at a distance of 2m gazed at seven positions in a prearranged random order, each 0.1m apart on a wall behind the participants. Participants made yes or no judgments of whether or not they felt that they were being looked at. The frequency distributions of 'yes' judgments showed a head-turn effect

such that when the target's head was rotated in one direction, participants' judgments tended to perceive gaze to be rotated in the opposite direction. In addition to the three head gaze conditions, Anstis et al. [4] investigated three orientations of a TV screen. They found three effects. First was a similar effect to the head-turn effect. Second was a TV-screen-turn effect where the apparent displacement of the perceived direction was in the same direction as the turn of the screen. Third was an overestimation of the deviation of the looker's gaze from the straight ahead. They suggested that the convex curvature of the screen probably caused the TV-screen-turn effect. Overestimation was found to increase with the complexity of the viewing condition. Overall, these studies suggest that observers may be constructing a mental line based on the head orientation before judging the eye direction relative to the head [78].

Despite the importance of the head as an attentional cue, there has been relatively little research on the perception of its orientation. Troje and Siebeck have provided evidence for the use of a head asymmetry cue to gaze [124]. Wilson et al. reported that head orientation discrimination is based upon both cues: deviation of head shape from bilateral symmetry, and deviation of nose orientation from vertical [133].

Perception of an avatar's gaze direction has also been studied in virtual environments [94, 115, 67]. Murray et al. [67] conducted three experiments to assess the efficacy of eye gaze within immersive virtual environments. The first experiment was conducted to assess the difference between users' abilities to judge what objects an avatar is looking at with only head gaze being viewed and also with eye- and head-gaze data being displayed. The results from the experiment show that eye gaze is of vital importance to the subjects, correctly identifying what a person is looking at in an immersive virtual environment. The second experiment tested subjects' ability to identify where an avatar was looking from their eye direction alone, or by eye direction combined with convergence. This experiment showed that convergence had a significant impact on the subjects ability to identify where the avatar was looking. The final experiment looked at the effects of stereo and mono-viewing of the scene, with the subjects being asked to identify where the avatar was looking. This experiment showed that there was no difference in the subjects ability to detect where the avatar was gazing. The authors also suggested several reasons why this may be the case. Firstly, the use of the chessboard for the avatar to look at creates an effective 3D effect due to other

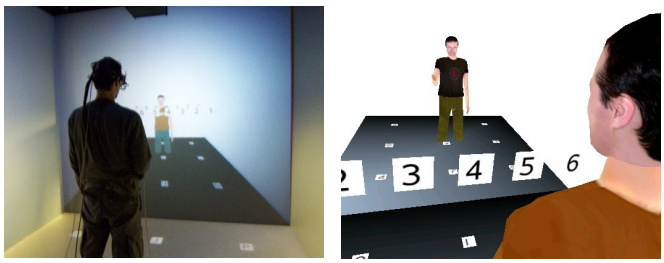


Figure 2.20: Example for evaluating gaze direction. [94]

depth cues, such as linear perspective and the chessboards texture gradient. Secondly, the subjects were located directly in front of the avatar.

Böcker et al. compared videoconferencing systems that provide motion parallax and stereoscopic displays and found this increased spatial presence and greater exploration of the scene [19]. Böcker et al. subsequently found the provision of motion parallax was shown to generate larger head movements in users of video conferencing systems [18]. Kim et al. found the combined presence of motion parallax and stereoscopic cues significantly improved the accuracy with which participants were able to assess gaze [50]. In Chapter 6, we further investigated the effects of reproducing motion parallax and stereoscopic cues in telepresence in both horizontal and vertical directions. We provided detailed reasons for the improvement of our system in conveying gaze.

In the first and second experiments (Chapter 4 and Chapter 5), we investigated the human gaze (both eye and head gaze) preservation capability of the spherical video telepresence system and cylindrical video telepresence system. In the third experiment (Chapter 6), which initially studies the use of random hole autostereoscopic multiview telepresence system for representing a remote participant, we employ the static gaze condition in evaluating random hole autostereoscopic multiview telepresence system, although the underlying system supports full eye gaze as well as facial expressions.

2.4.1.2 Evaluation framework

Detecting the gaze direction of a person is important for human computer interaction applications in video conferencing or shared collaborative workspaces. Evaluation of gaze includes object-focused gaze awareness and mutual gaze. Object-focused gaze awareness means that if the remote person is gazing at an object in the shared workspace, observers can know which the object is. Mutual gaze is the observers knowing whether remote person is looking at himself or herself. This is more com-

Table 2.3: Affordances of different telepresence systems

Display	Gaze	Group per site	Multi-user	360 view	3D
MAJIC [75]	√		√		
Hydra [110]	√		√		
GAZE-2 [128]	√		√		
MultiView [71]	√	√	√		
SphereAvatar [78]	√			√	
TeleHuman [50]	√			√	√
Spherical video system	√			√	
Spherical avatar system	√			√	
Cylindrical video system	√	√	√	√	
Random hole autostereoscopic multiview system	√	√	√		√

monly known as eye contact.

Nguyen et al. [71] proposed a framework for evaluation with three variables: attention source, attention target and observer. The attention source is a person who provides attention to the attention target. The attention target is an object which could be a person or anything else that receives attention from the source. The observer is the person who is trying to understand the presented information about attention including its source, its target, and any attached meaning.

In the object-focused gaze awareness situation, while a remote partner (attention source) fixed their gaze, the local participant (observer) was asked: Which object (attention target, such as numbered cards) is being looked at? In the mutual gaze situation, the local participant was asked: Are you being looked at? [94](see Figure 2.20)

2.4.1.3 Gaze in teleconferencing

Over the years a number of solutions have been developed to convey gaze direction during multiparty video conferencing, including MAJIC [75], Hydra [110], GAZE-2

[128], MultiView [71], animatronic shader lamps avatars [61] and One-to-Many System [46]. Also, a variety of solutions have been devised to explore the preservation of 3D depth cues and motion parallax via a single user head position tracking and the use of shutter glasses, such as, TeleHuman [50], SphereAvatar [78], 3-d live [87] and some CAVE-like environments [94, 40]. However, these systems are currently developed for a single observer. Table 2.3 summarised affordances of different telepresence systems. The last four display prototypes in this table are introduced by this thesis (Chapter 3).

2.4.2 Trust

Trust plays an important role in interpersonal communication. Sometimes, it is even an enabler for effective communication [65]. For example, in business settings, trust is required in order for coworkers or partner organizations to work together effectively. Without trust, partners will not share information openly, and transactions must be carefully contracted and monitored to prevent exploitation. Previous research shows that it can be more difficult to develop trust in teleconferencing than face-to-face [22, 72]. In this section, we first explore the relationship between trust and interpersonal cues: if interactions are mediated, some interpersonal cues are lost, thus more difficult to develop trust. We then look into the previous approach used in evaluating trust in telepresence systems. Lastly, we discuss the affordance of previous telepresence systems.

2.4.2.1 Trust and interpersonal cues

Trust can be defined as a 'willingness to be vulnerable, based on positive expectations about the actions of others' [65]. This suggests that trust is required in the presence of risk and uncertainty. Uncertainty arises from the fact that the user cannot directly observe the trustees ability (e.g. provider's skills, competencies, and expertise) and motivation (e.g. desire to deceive), but needs to infer those from cues [10]. Interpersonal cues can play an important role in the perception of trustworthiness in face to face situations, because they give information about an individuals background (e.g. education, provenance), but also about intrinsic states such as sincerity and confidence. Interpersonal cues include visual cues (e.g. appearance, facial expressions) and audio cues (para-verbal: e.g. pitch). However, these interpersonal cues can be lost in teleconferecing.

Table 2.4: The summary of social dilemma games for trust measurement

Name	Description
Prisoner's Dilemma [34]	Two men are arrested. They can choose either to defect or cooperate but without knowing the choice of the other. If one defects and the other cooperates, the betrayer goes free and the one that cooperate receives the full one-year sentence. If both cooperate, both are sentenced to only one month in jail for a minor charge. If each defects the other, each receives a three-month sentence.
Stag Hunt [55]	Two men go out on a hunt. They can choose to hunt a stage or a hare, but without knowing the choice of the other. If an individual hunts a stag, he must have the cooperation of his partner in order to succeed. An individual can get a hare by himself, but a hare is worth less than a stag.
Free Riding [27]	Comparing to two-person Prisoner's Dilemma task, this can be used to a larger group of individuals interacting with each other. Each person is better off using the bus without paying, but if everyone does this, the service will not be provided.
Daytrader [137]	Pairs of participants played a multi-trial variant of a Prisoner's Dilemma task, a task that has a long history of testing group cooperation and trust. Each participant was to imagine being a day-trader during a multi-day investment period.
Daytrader with Market Fluctuations [72]	These market fluctuations allow participants to withhold part of their investment and then blame the fluctuations for a lower than expected joint pay-off.
WindUp World [29]	Rather than just deciding on defection or cooperation, players navigated wind-up toys through a virtual world. When they met, the players had to decide whether they wanted to wind each other up (cooperation), or short-circuit the other player (defection).
Asynchronous Trust Game [10]	Unlike in the symmetric Prisoner's Dilemma game, the trustor first decides whether to trust the trustee or not.

2.4.2.2 Evaluation framework

As a measure of trust, a popular experimental paradigm currently employed by researchers has been social dilemma games, such as the Daytrader game [30]. Social dilemma games vary in how difficult they are depending on the exact rules and payoff structure, but it generally takes some amount of time and some communication in order to reach the required level of trust [22, 72, 89]. Those games are good models for synchronous and symmetric trust situations, such us two-way conversations. However, in some everyday trust situations, we can identify a trustor who decides first and a trustee who then decides to fulfil or defect, such as one-way conversation. Trust games can be suitable models of such situations [91]. Several social dilemma games for trust measurement are compared and summarised in Table 2.4.

Many researchers have investigated the relationship between trust and advice seeking behavior [12, 117]. Riegelsberger et al. [92] investigated users' trust in advisers and effects of media bias in different representations by observing participants' advice seeking behavior. In their scenario, participants were asked to participate in a quiz. Financial incentives were given for good performance. The questions included in the quiz were extremely difficult, so that good performance required seeking advice. Participants had two advisers but could only ask one for each question. Thus asking one adviser rather than the other can be understood as an indicator of trusting behavior. They found that users' preference for receiving video advice led them to disregard better text-only advice.

In chapter 7, we have followed the previous work [92] that has conceptualised trust in terms of individuals choice behaviour in a user-adviser relationship. We investigate two predictions regarding the effect of display type and viewing angle on trust: the spherical display may result in positive bias (i.e. more trust) because it increases social presence; or it may result in better discrimination between trustworthy and less trustworthy actors as it conveys more information.

2.4.2.3 The impact of eye gaze on trust in teleconferencing

Previous research indicates that it is hard to build trust in teleconferencing, because some non-verbal cues were unavailable to be 'read' [123]. To determine the effect of eye contact in video-mediated communication on trust, Bekkering and Shim [13]

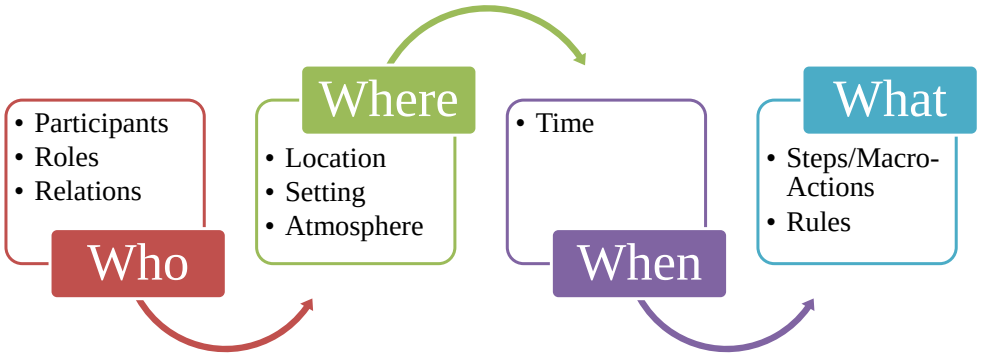


Figure 2.21: The framework for designing collaboration experience.

created a scenario in which participants indicated the trustworthiness of a message delivered by people. Results revealed that videos that did not support eye contact resulted in lower perceived trust scores, compared to videos that enabled eye contact. Voicemail enabled just as much trust as the video that created eye contact, perhaps because lack of eye contact cannot be perceived in audio-only communication. Nguyen and Canny [72] proposed a multiview video-conferencing system. They demonstrated that a video-conferencing system that affords more eye contact than the traditional video-conferencing system will create group trust levels similar to those seen in face-to-face group meetings.

Most of this previous work is focused on 2D planar displays. Trust formation on non-planar displays has not been evaluated yet. In chapter 7, we adapt previous studies of trust to evaluate the advantage of a sphere display over a flat display.

2.4.3 Designing collaboration experiences

Many scholars have proposed evaluation frameworks for most of the major application domains that telepresence systems could be used in [103, 93]. These application domains include communication actions, navigation and object-related actions. For communication actions, a sub-division differentiates between verbal (i.e., text and voice chat) and non-verbal communication (i.e., gestures, gaze, facial expressions, body posture, avatar appearance). The second category, navigation, comprises walking, flying, swimming, and teleporting. Object-related actions include the creation, selection, or insertion of objects [96]. Figure 2.21 illustrates the framework for designing collaboration experience.

We review two examples for designing collaboration experience. The first example

Figure 2.22: Examples of tiles participants were given to construct approximations of the logos. [42]

(a) Original situations for two users

(b) User A's progess

(c) User B's progess

(d) Guess together: BLACK BERRY

Figure 2.23: Tangram phase guessing game [21]

is to compare two methods of arranging multi parties distributed collaboration systems: around the table and same-side arrangement. Around the table, is when each user has a unique position and perspective, and any hand and arm gestures are seen by others to emanate from that position. However, for some tasks, such as reading a text, not all the users are capable of reading the text in the proper direction. The same-side arrangement where all collaborators see the table contents from the same perspective could solve this problem, but the participants are not able to see each other all the time. The tasks involved moving and arranging a set of tiles, initially piled in the centre of a shared workspace. Users were asked to recreate two of four possible logos (Figure 2.22, right) using a set of tiles containing various shapes (Figure 2.22, left). Users were required to use at least eight tiles and the tiles could be rotated and translated. This task mimics the photograph sorting or organization of many tabletop studies [42], where the content of the tiles and logos are less strongly oriented than the text-based task.

Another example is a comparison of competitive and cooperative task performance using spherical and flat displays [21]. An electronic variant of the Tangram game was chosen in this experiment. The original Tangram game requires users to arrange a number of different geometric pieces into various shapes within a silhouette of the target shape provided. This experiment modifies the basic Tangram game by combining it with a phrase-guessing game in order to increase the need for peeking and communication between participants, illustrated in Figure 2.23.

		Decision	
		Accept	Accept
		H0	H1
Reality	H0	Correct decision	Type 1 error
	H1	Type 2 error	Correct decision

Figure 2.24: Type of error in hypothesis testing according to the reality and the decision drawn from the test.

2.4.4 Statistical analysis

2.4.4.1 Hypothesis test procedure

In statistical data analysis an important objective is the capacity of making decision about population distributions and statistics based on samples. In order to make such decision a hypothesis is formulated abs tested using a appropriate mythology.

When we do a hypothesis test, two types of errors are possible: type 1 and type 2 (see Figure 2.24). The risks of these two errors are inversely related and determined by the level of significance and the power for the test. When the null hypothesis is true and we reject it, we will make a type 1 error. The probability of making a type 1 error is α, which is the level of significance we set for the hypothesis test. An α of 0.05 indicates that we are willing to accept a 5% chance that we are wrong when we reject the null hypothesis. To lower this risk, we must use a lower value for α. However, using a lower value for alpha means that we will be less likely to detect a true difference if one really exists. When the null hypothesis is false and we fail to reject it, we will make a type 2 error. The probability of making a type 2 error is β, which depends on the power of the test. We can decrease your risk of committing a type 2 error by ensuring our test has enough power. We can do this by ensuring our sample size is large enough to detect a practical difference when one truly exists.

2.4.4.2 Comparing means

There are different kinds of experiments design, namely, between subjects design, within subjects design and mixed design. Between group design is an experiment that has two or more groups of subjects each being tested by a different testing factor simul-

Table 2.5: The summary of ANOVA

Name	Description
One-way ANOVA	Provides a statistical test of whether or not the means of several groups, particularly more than two groups, are all equal.
Two-way ANOVA	Used for more than one independent variable and multiple observations for each independent variable. It can not only compare the main effect of contributions of each independent variable but can find out whether there is a significant interaction effect between the independent variables as well.
Repeated measures ANOVA	Used in repeated measure design and the repeated-measure factor is referred to as the within-subjects factor.
Mixed-design ANOVA	Used for two or more independent groups while subjecting participants to repeated measures. One factor is a between-subjects variable and the other is a within-subjects variable.

taneously. Within subjects design is an experiment in which the same group of subjects serves in more than one treatment. The mixed design is a combination of these two, which includes both between and within subjects variables.

Comparing to the between subject design, one of the greatest advantages of a within-subjects design is that it does not require a large pool of participants. Also, within-subjects design can help reduce errors associated with individual differences. However, a major drawback of using a within-subjects design is that the sheer act of having participants take part in one condition can impact performance or behaviour on all other conditions, a problem known as carryover effects. Additionally, fatigue is another potential drawback of using a within-subjects design. Participants may become exhausted, bored or simply disinterested after taking part in multiple treatments or tests.

For statistical analysis, there are different methods to analysis of variance (ANOVA) to find out variance in a particular variable partitioned into components attributable to different sources of variation, which are summarised in the Table 2.5. According to different experiment design methods, the Figure 2.25 shows how to select an appropriate analysing scheme.[45]

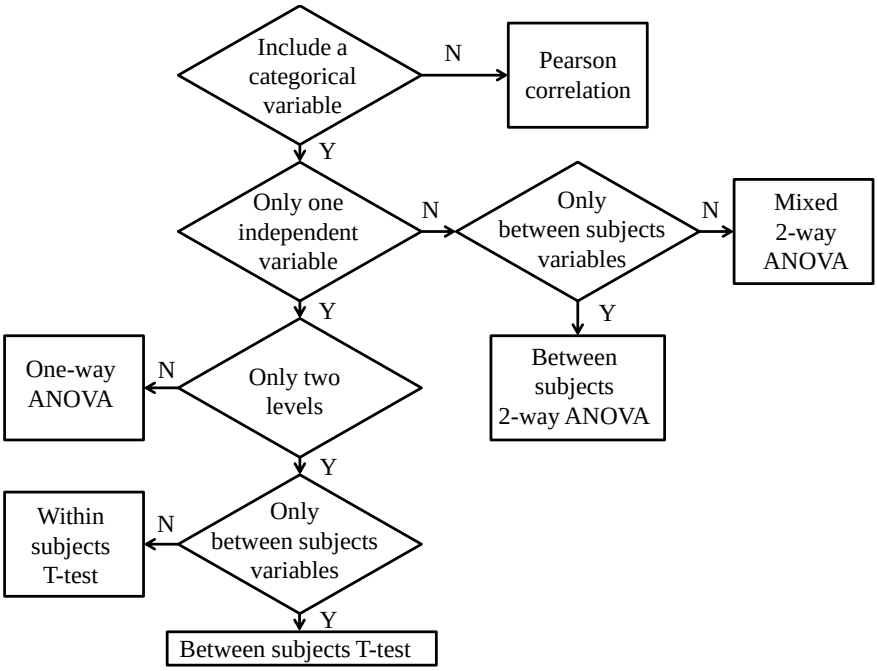


Figure 2.25: Flow chart to represent different choices of analysis experiment design.

2.5 Chapter summary

This chapter has been divided into four main sections. The first section discuss the important of nonverbal cues, detailed reasons for the gaze distortion in teleconferencing, and how gaze has been supported in different conversation scenarios, including two-way conversation, three-way or N-way conversation, group to group conversation and shoulder to shoulder conversation. In summary, for teleconferencing systems with a non-moving single observer, the impression of accurate gaze direction can be achieved through teleconferencing by aligning the camera through which a observer views a remote environment. For teleconferencing systems with multiple observers or a single observer at multiple viewpoints, the mona lisa effect occurs. For example, when a remote person looks into the camera, every observer seeing the video stream sees the remote person looking toward them. The reproduction of correct gaze direction is accomplished by providing unique and correct perspectives to each observer. Gaining inspiration from the previous teleconferencing systems which employ flat displays, the scope of this thesis concerned with situated telepresence systems and their affordances for one-way teleconferencing.

The second section presented an overview of teleconferencing display systems, covering situated display, multiview display, shape-changing display, virtual reality

system, augmented reality system and telepresence robot. Situated displays and multi-view display are covered in particular detail. The situated displays are visible from all directions, whereas flat displays are only visible from the front. These systems achieved maintaining accurate gaze by providing a perspective correct image via a single users head position tracking. Eventually only a mono or stereo image is presented on the display, thus they are currently developed for a single observer. The mutlview systems could support multiple users simultaneously each with their own perspective-correct view without the need for special eyewear. However, these are usually restricted to specific optimal viewing zones. The random hole display design which has a dense pattern of tiny, pseudo-randomly placed holes as an optical barrier mounted in front of a flat panel display. This allows observers anywhere in front of the display to see a different subset of the displays native pixels through the random-hole barrier. Building on previous research, we have built four telepresence system with different features, summarised in Table 2.3.

The third section introduced capture systems for both video and avatar mediated communication. For capturing perspective-correct videos, a remote person can be captured by a set of cameras and the video streams can be interpolated to achieve free viewpoint video. When the camera density is very high, view generation is simply by selecting the closest camera image. When the camera density is moderately high, view generation needs some processing. When the camera density is low, intermediate views can be generated by detecting geometry in the scene. In our spherical video telepresence system and cylindrical video telepresence system, the views are moderately dense, but we are not currently doing view interpolation. For avatar-mediated communication, where a remote person is represented by a graphical humanoid. Faceshift demonstrated a high-fidelity and real-time parametric reconstruction of facial expression method using a single depth camera. In spherical avatar telepresence system and random hole autostereoscopic multiview telepresence system, we have decided to represent a remote user as an avatar instead of video in our experiment, as 3D models are simple to render from any viewing angle.

The fourth section focused on the evaluation of teleconferencing systems. The affordance of object-focused gaze awareness and interpersonal trust in teleconferencing are discussed in detail. The particular framework of detecting the gaze direction of

a remote person was introduced. In this thesis, for the first three experiments (Chapter 4 to Chapter 6), we investigate the relationship between gaze and observer's viewing positions in different display configurations. Additionally, the particular scenario that conceptualised trust in terms of individuals choice behaviour in a user-adviser relationship was studied. For the last experiment (Chapter 7), we follow the previous work and investigate the influence of display type and viewing angle on how people place their trust during avatar-mediated interaction.

Chapter 3

System design

This chapter presents four novel telepresence systems that address research problems discussed in Section 1.2. The chapter details technical implementation to capture and display the remote person for the spherical video telepresence system, the spherical avatar telepresence system, the cylindrical video telepresence system, and the random hole autostereoscopic multiview telepresence system. Finally, the chapter presents a reading guide to the forthcoming experimental work.

3.1 Spherical video telepresence system

The goal of the spherical video telepresence system is to allow local users to perceive the eye gaze of a remote user accurately. Figure 3.1 depicts the system design. Table 3.1 presents the software and hardware components needed to implement our telepresence systems. A remote user, the *actor* in the *remote room* is captured by eleven capturing cameras controlled by two PCs. In the *local room*, a single PC renders video on a spherical display which is seen by a local user, the *observer*. Depending on the observer's position, the most appropriate camera feed is streamed from one of the two camera controller PCs to the renderer PC. Streaming is done using TCP. Table 3.2 shows the comparison of different network protocol to stream video.

3.1.1 Semicircular camera arrays

In the remote room, eleven low-cost PlayStation® Eye USB digital cameras are mounted on a half annular table with an inner radius of 405mm at every 15°, as illustrated in Figure 3.2. The cameras are set to the 56° field of view setting. The cameras capture at 30 Hz at 320×240 pixel resolution.

Table 3.1: Supporting tools of sphere display to presenting 3D real time video

Tools	Name	Feature and function
Hardware	Magic Planet Digital Globe	The digital display with a sphere-shaped screen
	16 Play station eye cameras	Capturing participant's head from different position
Software	CL-eye multi camera	Control multiple cameras, and select one or two video to stream
	Openframeworks	Client talk to server about which video should be selected via TCP protocol
	Open Graphic Library (OpenGL)	The environment for developing portable, interactive 2D and 3D graphics applications

Table 3.2: Comparison of different network protocol to stream video

Network protocol	Feature
Transmission Control Protocol (TCP)	Reliable services are able to ensure that packets are delivered to a host in the correct order. However, dropping packets is preferable to waiting for delayed packets via TCP, which may not be an option in a real-time system.
User Datagram Protocol (UDP)	A simple transmission model without implicit handshaking dialogues for providing reliability, ordering or data integrity.
Real-time Transport Protocol (RTP)	It is normally sent via UDP. It does not ensure "real time" but is a protocol that enhances the control and synchronization of real time video stream.
Real-time Streaming protocol (RTSP)	Control multiple data delivery sessions, provide a mean for choosing delivery channels such as UDP, multicast UDP and TCP, and provide a means for choosing delivery mechanisms based upon RTP.
Hyper Text Transfer Protocol (HTTP)	Streaming video can be sent via HTTP "tunneling", since virtually all firewalls allow the default http port (port 80) to pass. There is a severe penalty, HTTP is sent via TCP which increases the overhead by some 30% and magnifies the delay.

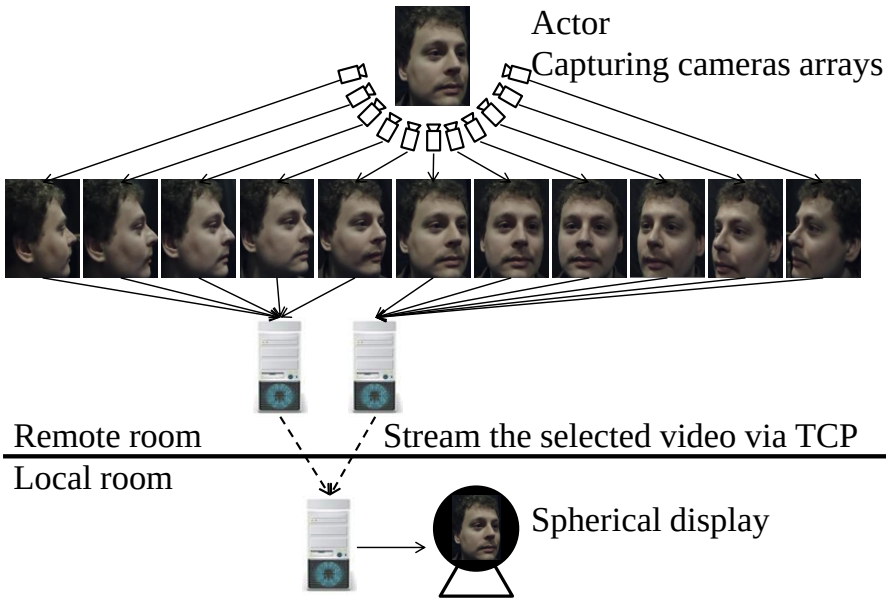


Figure 3.1: Diagram of the directional spherical video conferencing system.

Figure 3.2: Camera calibration setup.

We manually adjust the cameras to look at the point above the centre of the half annular table. We use Zhang's camera calibration method which involves showing all of the cameras a planar checkerboard target in at least two different orientations [136]. We then use Camera Calibration Toolbox for Matlab® to locate the cameras' positions and orientations accurately (in Figure 3.2). These positions and orientations are used in the rendering process.

3.1.2 Directional spherical screen

In the local room, a spherical display is located at the centre of a half annular table which is the same size as the one in the remote room. Eleven observer viewpoints set around the half annular table with a radius of 1810mm at every 15° which exactly line

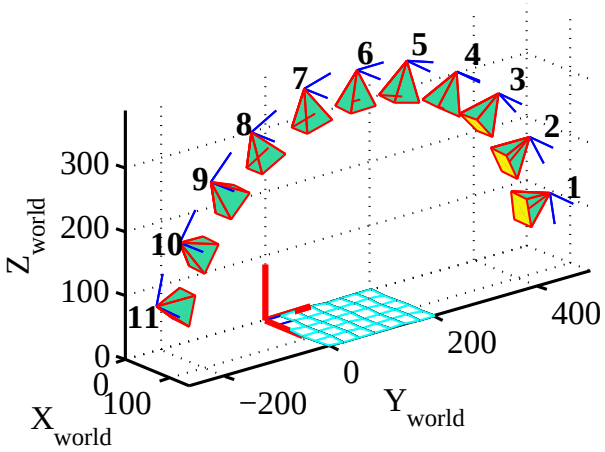
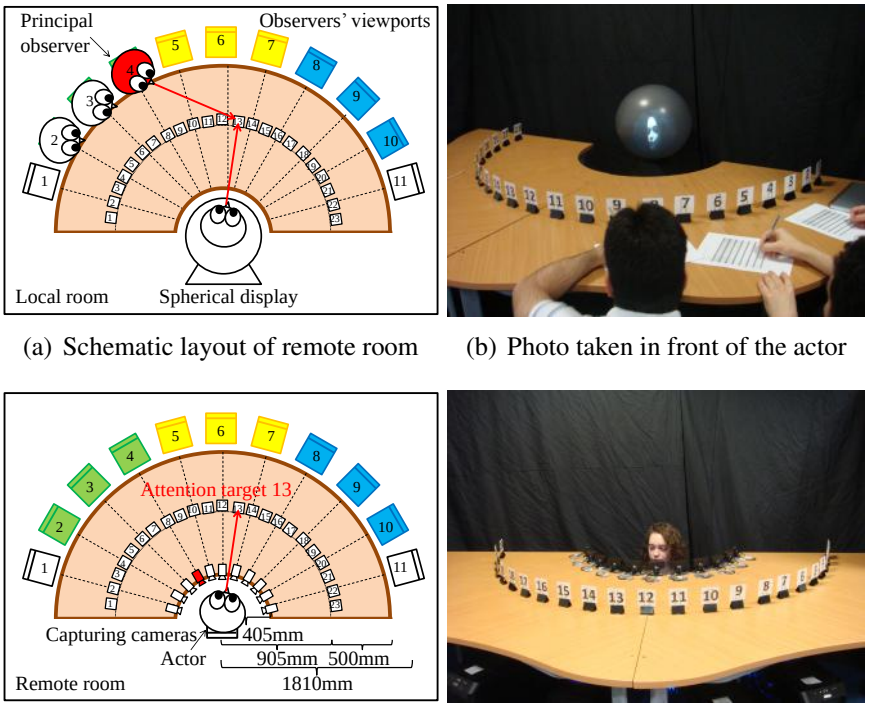


Figure 3.3: Camera calibration result.



(a) Schematic layout of remote room (b) Photo taken in front of the actor

(c) Schematic layout of local room (d) Photo taken behind observer 4

Figure 3.4: Example of system & experiment setup: The actor gazes at the target card 13 captured by semicircular camera arrays in remote room. Since the principal observer is seating in viewpoint 4, the video captured by camera 4 is presented on the sphere display, which lines up with the observer 4.

up with each camera in the remote room as depicted in Figure 3.4(c). The spherical display is the commercially available Magic Planet display by Global Imagination®. The Magic Planet is a projection display device with a 16" sphere-shaped surface and an internal fisheye lens to project imagery on to the inside of the sphere.

The presentation of the remote participant onto the sphere is done in four main

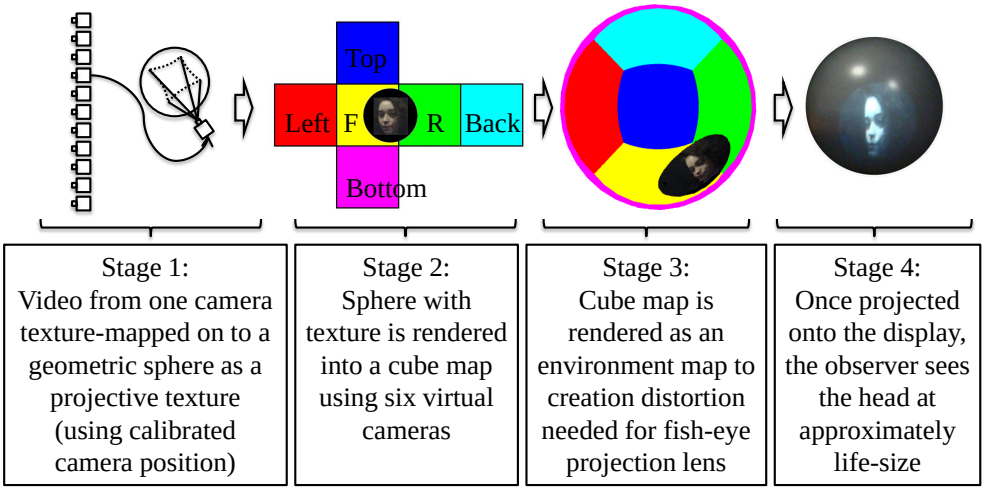
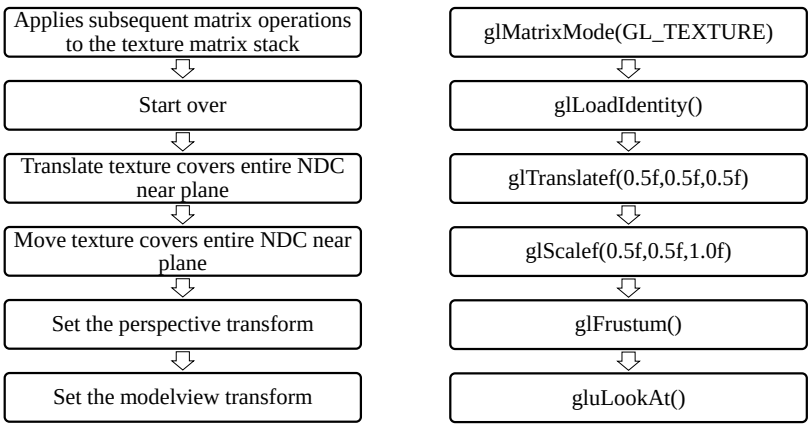


Figure 3.5: Illustrating stages of the rendering pipeline. Note: In the cube map and the 2D distorted image, the coloured background representing six different faces of a cube is just for the sake of explanation. Actually, it is all black.



(a) algorithm (b) OpenGL functions

Figure 3.6: Flow chart of projective texture

stages shown in Figure 3.5.

First, a sphere acted as a proxy geometry of a human head, on to which the video images are displayed using projected texture mapping (PTM). PTM is a method of texture mapping described by Segal that allows the texture image to be projected onto the scene as if by a "slide projector" [107]. Figure 3.6 shows the flow chart of implementing the projective texture. According to the observer's viewpoint, the video captured by a corresponding capturing camera is selected. This video is projected onto the polyhedron, which is approximately human head size. This ensures that the capturing camera, the "slide projector" and the observer's eye are in close alignment.

Next, we rendered this proxy geometry in to an environment map. The idea of storing environment maps as the cube maps is proposed by Greene where six subimages representing the six different faces of a cube [39]. We rendered the scene in to an environment map using six cameras positioned outside the cube at the position of the observer's eye. Each of the six facets of cube map is thus rendered using the non-symmetric view volumes. The resulting cube map looks as if the head is outside looking in, but once reflected in the environment mapping, it gives the illusion that the head is situated within the spherical display.

Then, we draw a 3D sphere using an environment map. Environment mapping proposed by Blinn and Newell simulates the reflectance of a surface, by using the reflected eye vector as a lookup in to the texture rather than a simple texture coordinate [16]. We render a sphere with the environment map as its texture in order to generate a 2D distorted image, that is suitable for projection through a fish eye lens [78].

Finally, the projected light travels through the bottom of the sphere, allowing the sphere to be completely illuminated except for the area immediately around the lens itself and achieving 360° horizontal visibility. The observer sees the head life-size.

3.2 Spherical avatar telepresence system

The spherical avatar telepresence system captured a remote person's interpersonal cues and represented them as an animated avatar head on a spherical display. In the remote room, the facial expression of the remote person, the *actor*, is captured. In the local room, a single PC renders an animated avatar on a spherical display which is seen by an observer. Figure 3.7 depicts the system design. We integrated with Faceshift® to allow an actor to control the facial expressions of the avatar. We developed a view-dependent (depending on observers' viewing positions) graphical representation to fully support rendering spherical display surfaces.

3.2.1 Real time facial expression tracking with Faceshift

In the remote room, the actor is recorded in a natural environment using a non-intrusive, commercially available Microsoft Kinect™. The actor was seated at the same height as the sensor, about 600 mm horizontal distance from the sensor (see Figure 3.7). The Microsoft Kinect™ supports simultaneous capture of a 640×400 2D color image and a

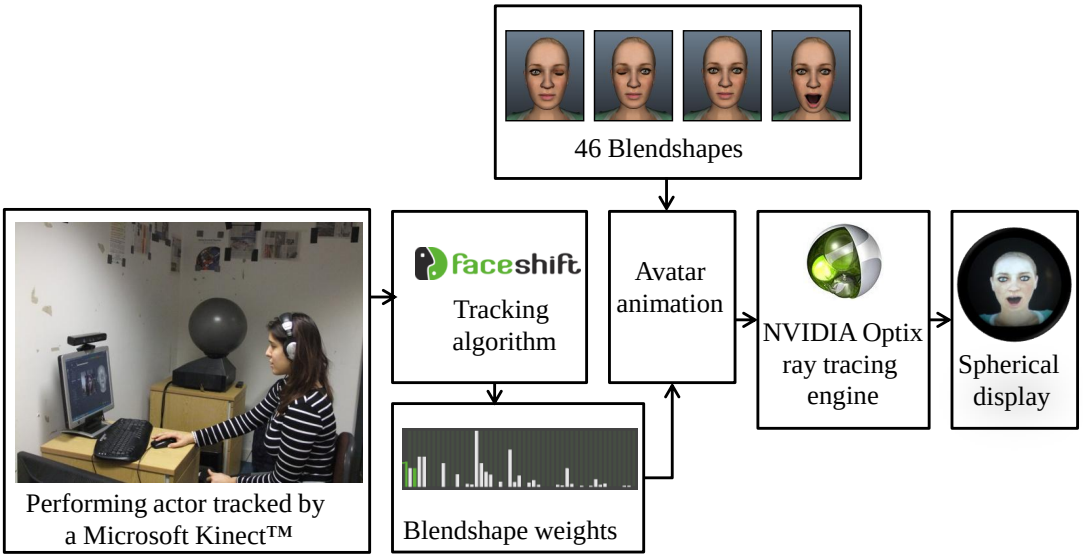


Figure 3.7: Pipeline for representing an avatar with dynamic facial expressions controlled by an actor on the spherical display.

3D depth map at 30 Hertz, based on invisible infrared projection. It provides a simple and low cost way for acquisition, without the use of face markers, intrusive lighting, or complex scanning hardware.

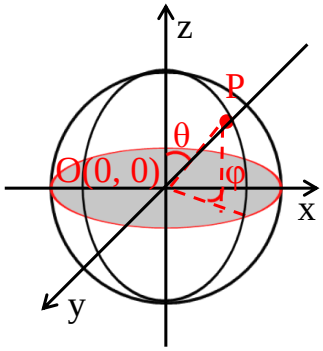
We used Faceshift® with Microsoft Kinect™ to obtain our actor's facial performances in realtime. Faceshift® ensures robust processing given the low resolution and high noise levels of the input data. The output of the tracking optimization is a continuous stream of blendshape weight vectors that drive the avatar. With the embedded plug-in of Faceshift® in Maya®, we obtained 46 blendshapes of the Rocketbox® avatar by Maya® then exported them as .obj format for the usage of ray tracing stage discussed in next subsection. Finally, we represented facial expressions as a weighted sum of blendshape meshes, enabling actor to control the facial expressions of the avatar.

3.2.2 View dependent rendering for spherical display

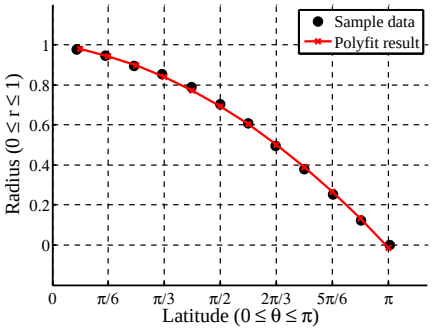
In the local room, we used the same commercially available spherical display as the spherical video telepresence system discussed above (see Section 3.1.2).

We developed a view dependent rendering method to create 3D object presenting onto spherical image surface, map from the spherical image surface into 2D image plane, and re-project onto spherical display, as if the object is situated inside the sphere display (See Figure 3.8).

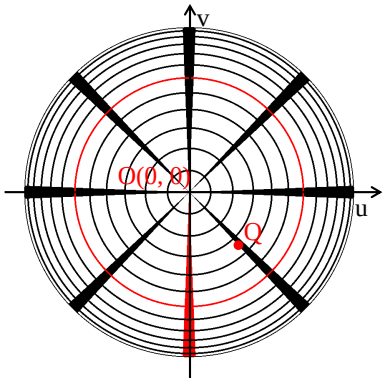
We used the NVIDIA® OptiX ray tracing engine [83]. We traced the path of light



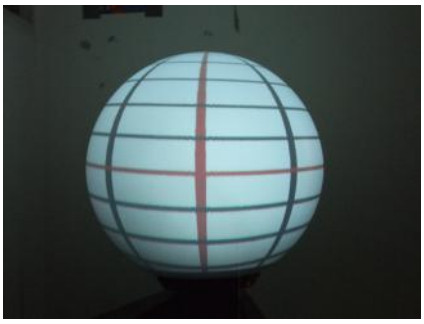
(a) 3D Sphere image surface



(b) The relationship between latitude angle (θ) and the radius (r)



(c) 2D mapping image plane generated for projection



(d) Photo taken in front of the sphere display

Figure 3.8: The mapping relationship: each point P on the 3D spherical surface in the subfigure (a) translates into corresponding point Q on the 2D image plane in the subfigure (c), according to calibrated relationship in the subfigure (b). The subfigure (d) shows the projected result of the 2D image plane.

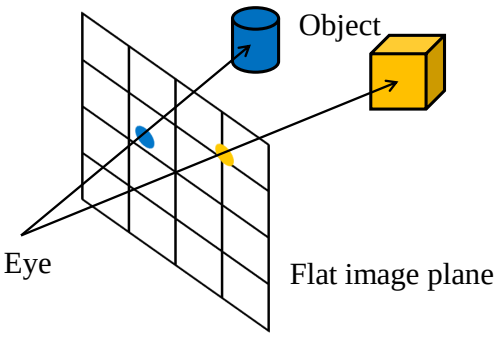


Figure 3.9: Flat image plane ray tracing.

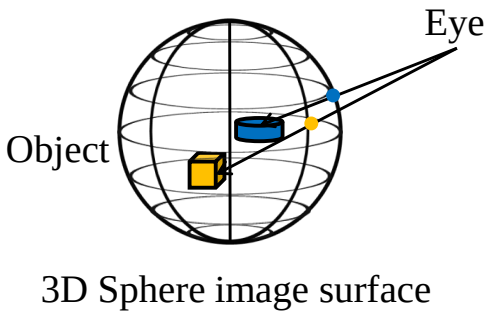
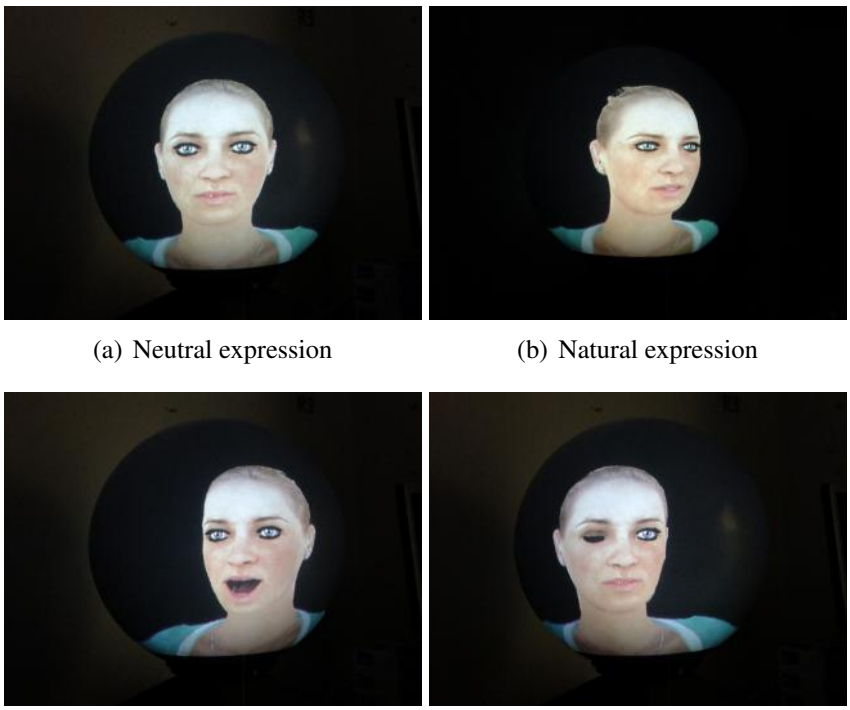


Figure 3.10: Spherical image surface ray tracing.

(a) 0° (b) 15° (c) 30° (d) 45° (e) 60° (f) 75° (g) 90°

Figure 3.11: 2D mapping image generated for projection at different viewer positions.



(a) Neutral expression (b) Natural expression

(c) Mouth open (d) Left eye close

Figure 3.12: Photo taken at approximately 45° left side of sphere display. For both subfigure (a) and (b), the viewers' positions are the same as the photo taken position. The avatar head is looking at the right of the viewer in the subfigure (a), but the avatar head is looking at the right of the viewer in the subfigure (b). For subfigure (c) and (d), each viewer's position is at right and left side of the photo taken position, respectively.

from observer's eye to the 3D object though pixels in a spherical image surface. Ray tracing is a computer graphic technique to generate an image by tracing the path of light through pixels in an image plane and simulating the effects of its encounters with virtual objects [38][111]. Figure 3.9 represents the traditional ray tracing technique to create 3D world to 2D image plane. We use a similar idea by tracing the path of light though pixels in a spherical image surface. This is illustrated in the Figure 3.10. The use of a ray tracing engine should provide higher quality images with less distortion than the polygonal rendering approach that was developed for SphereAvatar [78].

To implement the ray tracer, we translate the 3D spherical surface into 2D image plane, to represent the surface of the sphere on a flat paper map or on a computer screen. The position of each point (Q) on the 2D image plane (see Figure 3.8(c)) can be defined by a radius $(r,\ 0 \leq r \leq 1)$ and a longitude angle $(\alpha,\ 0 \leq \alpha \leq 2\pi)$. The corresponding position of that point (P) on the spherical surface (see Figure 3.8(a)) can be defined by a latitude angle $(\theta,\ 0 \leq \theta \leq \pi)$ and longitude angle $(\phi,\ 0 \leq \phi \leq 2\pi)$. The 2D image projector and the display surface are axially symmetric about the optical axis. Thus, the polar angle (α) in 2D image plane is the same as the longitude angle (ϕ) in the 3D spherical surface, shown the Equation 3.1. All the points at a given radius (r) in the 2D image plane are projected onto the sphere display surface at the same latitude angle (θ). Because of lens distortion, there is a nonlinear relationship between latitude angle (θ) of sphere display surface and the radius (r) of the 2D image plane. We sampled latitude angle (θ) at every 15° to find out the corresponding radius (r) value, shown in Figure 3.8(b). We used the Matlab® second order polyfit to simulate a continuous function as a model to characterize the relationship between the latitude angle (θ) and the radius (r), presented in Equation 3.2. Therefore, if we want to project a certain image onto sphere display surface, the corresponding source image can be determined by applying the inverse function to that image.

$$\alpha = \phi \tag{3.1}$$

$$r = -0.0806 \times \theta^2 - 0.0704 \times \theta + 1.0022 \tag{3.2}$$

Finally, the 2D image plane produced would be projected through the fisheye lens of the sphere display. We could then see a corrected image presented on the spherical

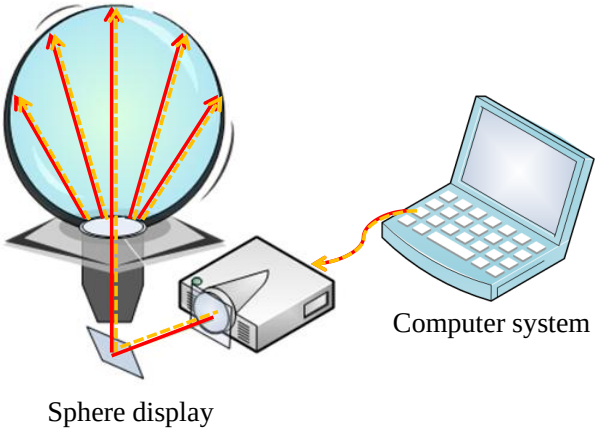


Figure 3.13: A stereoscopic representation on sphere display.

surface. In Figure 3.8(c), the red circle of the 2D image plane is corresponding to the equator of the sphere; the center of the 2D image plane projects to a single point on the top of the sphere; the very outer circle in the 2D image plane projects to a single point on the bottom of the sphere. The projected result on sphere display is presented in Figure 3.8(d).

We use this view dependent graphic representation method discussed above to ray trace an avatar's upper body and head purchased from Rocketbox®. Figure 3.11(a) to Figure 3.11(g) present some sampled mapping results in 2D image plane generated at different viewers positions while the avatar is looking at the front. Once projected through the fisheye lens of the display, such images would appear as correctly shaped head and upper body (see Figure 3.12).

This method successfully avoids any seams, overlaps or registration errors in the resulting composite image in projecting image on sphere display. It also could extend to other display systems that have a three dimensional display surface.

We could use a tracking device to obtain the viewer's position and view direction. Then, we could obtain the image with correct view present on sphere display at 360° free viewpoint positions.

We also could produce 3D effect on the sphere display if the viewer is in a fixed position or wearing an eye tracking device. We could use the techniques described above to create two different desired image for each eye. Then, the resulting flat source images are output for sphere display by a the stereoscope projector. The viewer could see the 3D effect by wearing the eye-gear appropriate for the projector, such as polar-

Qty	Item	Cost/Unit	Total cost
1	Retro-reflective Sheet	£10	£10
1	Lenticular Sheet	£20	£20
4	Camera	£20	£80
4	Projector	£350	£1400

Table 3.3: Cost for a set up for four observers.

ized glasses for a polarizing stereoscope projector (see Figure 3.13).

3.3 Cylindrical video telepresence system

The goal of the cylindrical video telepresence system is to allow multiple observers to perceive the gaze of a remote person accurately. That is observers can each see a unique and perspective-correct image from their viewing directions simultaneously. The spherical video telepresence system (see section 3.1) and spherical avatar telepresence system (see section 3.2) only supports a single observer.

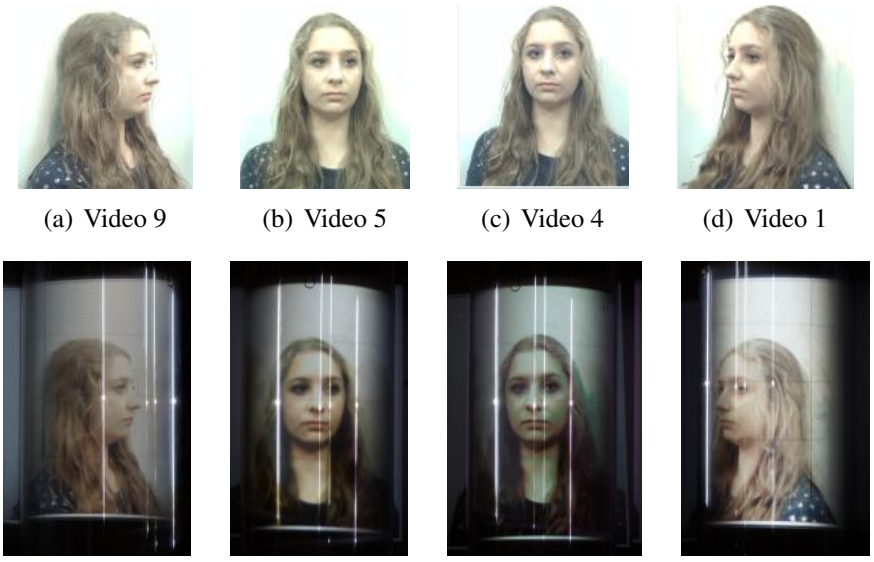
In this system, each camera is linked to the corresponding projector to stream real-time video using TCP. The cylindrical screen ensures that each projected image will only be seen by an observer who is in the viewing zone for that projector. Also, using available off-the-shelf components allows our system to be built at a low cost (see Table 3.3).

3.3.1 Semicircular camera array construction

In the remote room, the capture system of this system was similar to the spherical video telepresence system. Nine PlayStation® Eye USB digital cameras were vertically mounted on an angled table at a radius of 600mm every 15°, as illustrated in Figure 3.15(a) and Figure 3.15(b). We manually adjusted the cameras to look at the point above the center of the angled table. We then used Camera Calibration Toolbox for Matlab® to locate the cameras' positions and orientations accurately. The accurate positions and orientations of the cameras are used in the arrangement of projectors, so that accurate projecting of video can be done. The cameras were set to the 56° field of view setting. The cameras capture at 30 Hz at 640×480 pixel resolution. We arranged cameras vertically in order to make full use of the pixel resolution to represent the remote person's head.

Table 3.4: Summary of materials for multiple layers of the screen design.

Methods	Material	Function	Product
The diffused retro reflector method, utilized a retro reflector to return the light in the direction of the projector for horizontal retro reflection, and diffusion layer for vertical diffusion.	Retro reflect layer	Provide a strong retro reflective specification. An ideal retro reflective material bounces all of the light back to its source. (see Figure 3.16)	Chromatte; "white number plate reflective"
	One-dimensional diffuser layer	A lenticular sheet was used as the diffuser. A spacing of 1/4" or more between retro reflect and lenticular sheet is recommended, otherwise the diffusion effects of the lenticular will be undone by the retro reflect.	The lenticular sheet with 40 lenticules per inch
	Anti-glare layer	The high gloss finish of the lenticular sheets produced a very distracting glare along the path of reflection.	Grafix Matte (Frosted) Acetate
The lenticular method, utilized lenticular image for horizontal retro reflection and vertical diffusion.	Lenticular lens	A lenticular sheet was used for horizontal retro reflection and vertical diffusion. (see Figure 3.17)	MicroLens, Lenticules per Inch: 30
	Diffusive backing	The first sheet was affixed to the lens using the adhesive on the lenticular lens. The second sheet was affixed to the first sheet of paper using a spray adhesive.	Two sheets of ink jet plotter paper.



(a) Video 9 (b) Video 5 (c) Video 4 (d) Video 1

(e) View 9 (f) View 5 (g) View 4 (h) View 1

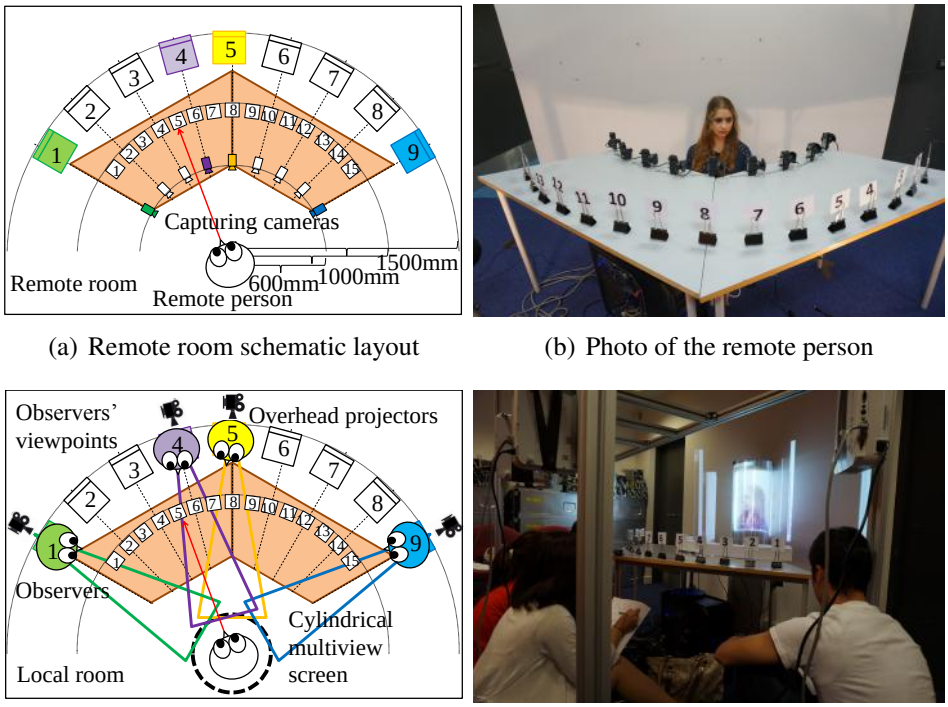
Figure 3.14: The top row is four videos simultaneously captured from four different cameras. The bottom row is four photos of the same display from four different perspectives. The remote person is gazing at target 5. See Figure 3.15 for camera, target and viewpoint numbers.

3.3.2 Cylindrical multiview screen design

In the local room, the cylindrical screen is located at the center of an angled table which is the same size as the one in the remote room. We designed a cylindrical screen 32 cm in diameter and 70cm in height. The size is small enough to situate almost anywhere in a room. This display is visible from all directions, whereas flat displays are only visible from the front. The radius of curvature of the screen is similar to a real convex face to avoid the TV-screen-turn effect [4]. Using a cylindrical screen surface significantly simplifies the projection of correct vertical perspective to observers at different heights and distances from the display.

The screen's main function is to reflect the image produced by a projector only to an observer in a very specific viewing zone. This could be achieved based on diffused retro reflector method and lenticular methods. The material needed in both diffused retro reflector method and lenticular methods are summarised in the Table 3.4. We present the detailed design in the Figure 3.18. The "Front View" shows a small amount of diffusion in the horizontal directions. The "Side View" shows a large amount of diffusion in the vertical direction.

The screen consists of a retroreflective layer around the cylinder, with a one-



(a) Remote room schematic layout

(b) Photo of the remote person

(c) Local room schematic layout

(d) Photo taken behind observer 1

Figure 3.15: Experiment setup: In the remote room, a camera array is used to capture unique and correct perspectives of the remote person gazing at the target 5. In the local room, a cylindrical multiview display is used to allow each observer to view their respective perspectives simultaneously. One of observers seating in viewpoint 1, only sees the video captured by camera 1.

dimensional diffuser layer 6mm above. Experimentation was conducted with different retroreflective materials, leading to the decision to use "white number plate reflective" from ORALITE®, because it has a strong retroreflective characteristic, minimal reflective properties and good diffusive properties to reduce glare effects. A 1D lenslets-based lenticular sheet was used as the one-dimensional diffuser. The lines of the lenticular sheet were placed horizontally to provide vertical diffusion. A 6mm or more physical spacing between retroreflective layer and lenticular sheet allowed the light to mix vertically. The smooth side of the lenticular sheet was facing the observers and projectors. The 40 lenticules per inch (LPI) sheet with 49° viewing angle from Pacur® was chosen for two reasons: the thin thickness (0.838mm) of this sheet allowed it easily to wrap around the cylinder; and we only require a modest amount of vertical diffusion. More diffusion would lower the brightness of the image. The screen's optics carefully retro reflects the light in the direction of the projector but diffuses it vertically,

(a) Perfect mirror (b) perfect retro reflector

Figure 3.16: A comparison of reflection and retro reflection

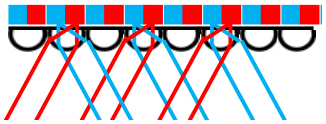
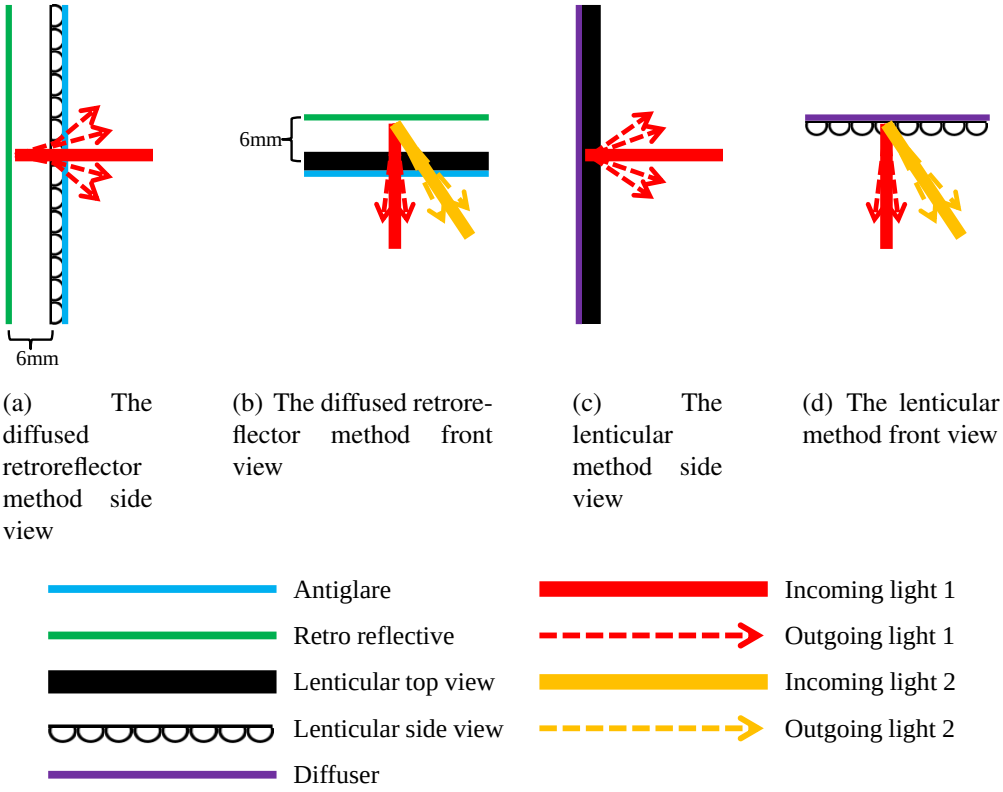


Figure 3.17: An example of using the lenticular method as a front-projection multiview screen.



(a) The diffused retroreflector method side view

(b) The diffused retroreflector method front view

(c) The lenticular method side view

(d) The lenticular method front view

Figure 3.18: Multiple layers of the screen design.

allowing viewers to see the image from any position above or below the projector.

3.3.3 Semicircular projector arrays construction

Nine projectors and observer viewpoints were set around the half annular table with a radius of 1500mm at every 15° which exactly line up with each camera in the remote room as depicted in Figure 3.15(c) and Figure 3.15(d). We vertically mounted each projector at a height of 1800mm, allowing a observer to sit under a projector. We used Projector-Camera Calibration Toolbox® to align the projectors' positions and orientations accurately. Each projector projected a unique image on the part of the cylinder at the same horizontal level, but there were some overlaps between images that are projected by different projectors. The cylindrical multiview screen controls diffusion and produces relatively narrow viewing zones above, below, and slightly to the sides of a light source. Therefore, an observer sitting under the bottom of a projector sees only the image from that projector. We used NEC® NP110 projectors with resolutions of 800×600 pixels.

3.4 Random hole multiview telepresence system

The use of autostereoscopic display technologies could support multiple users simultaneously each with their own perspective-correct view without the need for special eyewear. However, these are usually restricted to specific optimal viewing zones.

Our telepresence system uses the random hole display design [134, 69] which has a dense pattern of tiny, pseudo-randomly placed holes as an optical barrier mounted in front of a flat panel display. This allows observers anywhere in front of the display to see a different subset of the display's native pixels through the random-hole barrier. Additionally, it is technically quite simple to build and can be constructed cheaply in comparison to holographic and volumetric displays.

We developed a view-dependent ray traced rendering method to represent a remote person as an avatar on the random hole display. The method allows multiple observers in arbitrary locations to perceive stereo images simultaneously.

3.4.1 Hardware

Our hardware is based on the design of Gu et al.'s Tabletop Autostereoscopic Display [134]. The display used three layers to create its viewing zones. A diagram of the

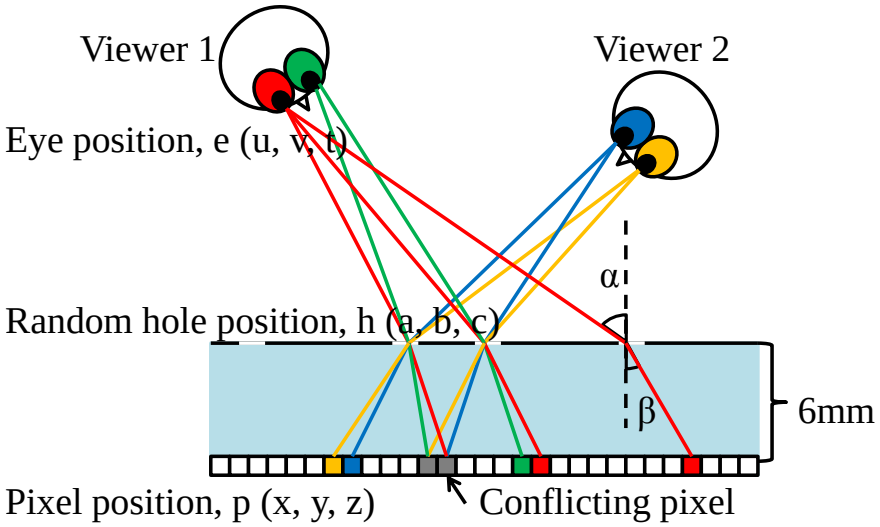


Figure 3.19: A top down diagram of the random hole display showing two viewing positions.

layers is shown in Figure 3.19. The back-most layer is a single LCD display panel. The HP ZR30w 30-inch S-IPS LCD Monitor was used for two reasons. Firstly, as a parallax barrier reduces the effective resolution of the display, we selected a high-resolution (2560×1600) and reasonably priced LCD. Secondly, we used the S-IPS type display, because it has very large horizontal and vertical viewing angles. In contrast, the twisted-nematic (TN) panels, which are widely used for low-cost consumer-grade LCD displays, have a limited vertical viewing angle and exhibit colour inversion when viewed from below. The next layer is a Lexan™ polycarbonate sheet, which forms the separating layer. The thickness of the sheet is 6mm (approximately £30). The Lexan™ polycarbonate sheet's refractive index is slightly above 1.5 and similar to the index of the LCD panel's built-in transparent cover. The last layer is the random hole mask that was printed on a thin polyester film at 1200 dpi (approximately £15).

3.4.2 Software

We developed a view-dependent ray trace rendering method to represent a remote person as an avatar on the display. We used the NVIDIA® OptiX ray tracing engine. Instead of tracing a ray from a viewpoint through each pixel in a virtual screen, we traced a ray from each eye through each hole in the mask (see Figure 3.19).

Each eye position, each random hole position and each pixel position can be defined as $e(u,v,t)$, $h(a,b,c)$ and $p(x,y,z)$ in cartesian coordinates. The angles of incidence can be defined by a latitude angle (θ_1, $0 \leq \theta_1 \leq \pi$) and longitude angle

Data: The position of each eye and the position of each hole in the mask
Result: For each pixel in the screen, store the origin and the direction of its corresponding ray
Initialize *screen* of size $h \times w$ to zero;
Set $conflicting\_count$ of the corresponding pixel position to zero;
foreach *eye in the eye position* **do**
  foreach *hole in the hole position* **do**
    Assign $ray\_origin = eye$;
    Calculate $ray\_direction = eye - hole$;
    Calculate the corresponding pixel position hit by *ray* through *hole* based on Snell's law;
    if $conflicting\_count > 0$ **then**
      Choose one of the conflicting view randomly;
    end
    Store ray_origin and $ray\_direction$ in the corresponding pixel position of *screen*;
    Add 1 to $conflicting\_count$;
  end
end

Algorithm 1: Store the corresponding ray for each pixel.

($\phi_1,\ 0 \leq \phi_1 \leq 2\pi$) in geographic coordinates. The corresponding angle of refraction can be defined by a latitude angle ($\theta_2,\ 0 \leq \theta_2 \leq \pi$) and longitude angle ($\phi_2,\ 0 \leq \phi_2 \leq 2\pi$) in geographic coordinates.

The angle of incidence could be computed based on the relationship between the cartesian coordinates and geographic coordinates, presented in Equation 3.3c and Equation 3.4, where r_1 is the distance between the eye position to the hole position.

$$u - a = r_1 \cdot \sin\theta_1 \cdot \cos\phi_1 \tag{3.3a}$$

$$v - b = r_1 \cdot \sin\theta_1 \cdot \sin\phi_1 \tag{3.3b}$$

$$t - c = r_1 \cdot \cos\theta_1 \tag{3.3c}$$

$$r_1 = \sqrt{(u-a)^2 + (v-b)^2 + (t-c)^2} \tag{3.4a}$$

$$\theta_1 = \arccos(\frac{(t-c)}{r_1}) \tag{3.4b}$$

$$\phi_1 = \arctan(\frac{v-b}{u-a}) \tag{3.4c}$$

We consider the refractive effects when the light passes through the barrier film, the Lexan™ polycarbonate sheet separating layer, and the LCD panel's own protective cover.

We compute the angle of refraction according to Snell's law that the ratio of the sines of the angles of incidence and refraction is equivalent to the reciprocal of the ratio of the indices of refraction (see Equation 3.5).

$$\theta_2 = \arcsin(\frac{\sin\theta_1}{n}) \tag{3.5a}$$

$$\phi_2 = \phi_1 \tag{3.5b}$$

We then find the corresponding pixel position, based on the relationship between the cartesian coordinates and geographic coordinates, presented in Equation 3.6, where r_2 is the distance between the pixel position to the hole position.

$$x = r_2 \cdot \sin\theta_2 \cdot \cos\phi_2 + a \tag{3.6a}$$

$$y = r_2 \cdot \sin\theta_2 \cdot \sin\phi_2 + b \tag{3.6b}$$

$$z = r_2 \cdot \cos\theta_2 + c \tag{3.6c}$$

Next, we calculated the color of the object visible on a certain area of the screen through each hole for each eye. If multiple eyes see the same pixels behind the barrier, then a conflict occurs. We chose the color from one of the conflict views randomly. By using the pseudo-random Poisson distribution of the hole pattern [33], visual conflicts

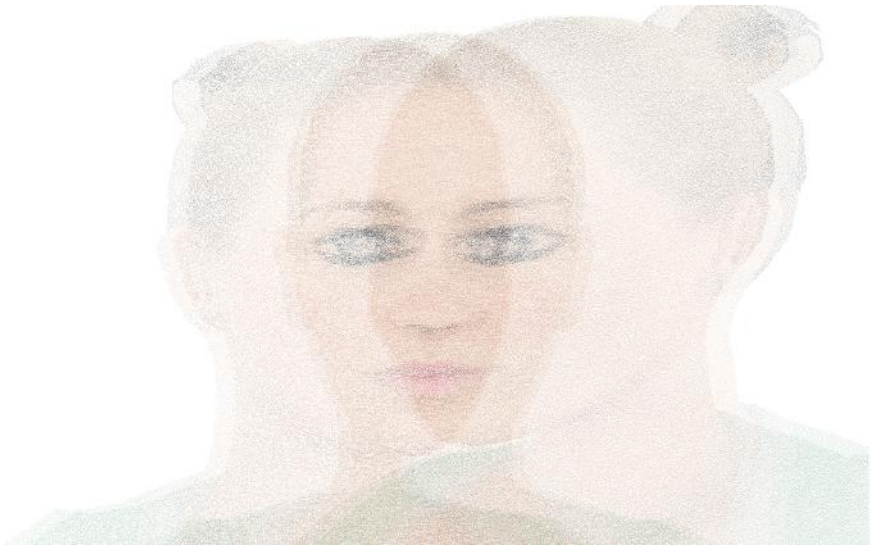


Figure 3.20: Source image of six simultaneous views

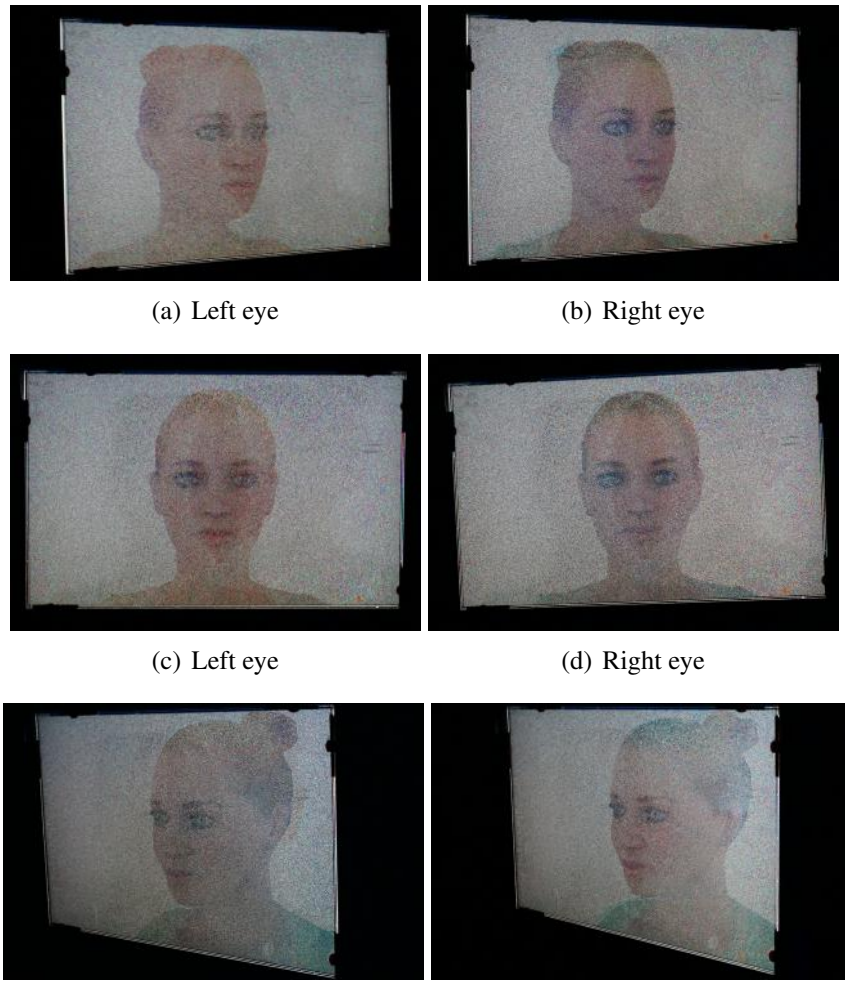
between views are distributed across the viewing area as high frequency noise. The high frequency noise is typical of these displays; however, users can clearly identify images and objects.

Figure 3.20 shows the source image (six views) actually displayed on the LCD panel. It allows three observers in front of the display to see perspective-correct stereo images on the subset of the display's native pixels through the random-hole screen. Figure 3.21 shows photographs from six viewing positions, corresponding to the three stereo views of the three observers.

3.5 Chapter summary

This chapter presented the design and implementation of our four teleconferencing systems which could be used in telepresence applications. The spherical video telepresence system and the spherical avatar telepresence system support perspective correct view for a single observer at multiple viewpoints. The cylindrical video telepresence system extended this capability to multiple observers at multiple viewpoints. The random hole autostereoscopic multiview telepresence system further improved the affordance of teleconferencing experiences, by adding stereoscopy cues.

Table 3.5 is designed as a reading guide to the upcoming experimental chapters. The three evaluations on the affordance of object focused gaze in telepresence systems are reported in chronological order, with increasing capability regarding the performance of telepresence systems. This is followed by one evaluation on the affordance of



(a) Left eye (b) Right eye

(c) Left eye (d) Right eye

(e) Left eye (f) Right eye

Figure 3.21: Photos of six simultaneous views of the random hole display at 170cm from the display.

interpersonal trust in a telepresence system. The chapters in which each may be found is shown in the rightmost column. The system column refers to the teleconferencing systems used in the experiment. The media refer to the video or avatar content used in representing the remote person. The column headed affordance indicated the capability of the teleconferencing systems in terms of gaze-preserving or interpersonal trust.

Table 3.5: Overview of experimental chapters. For each chapter we list: telepresence systems used, the media used in communication, and evaluations on the affordance of the telepresence systems.

Chapter	System	Media	Evaluation
Chapter 4	Spherical video telepresence system	Video	Gaze-preserving capability for a single observer at multiple viewpoints
Chapter 5	Cylindrical video telepresence system	Video	Gaze-preserving capability for multiple observers at multiple viewpoints
Chapter 6	Random hole autostereoscopic multiview telepresence system	Avatar	Gaze-preserving capability for multiple observers at multiple viewpoints, augmented by stereoscopy.
Chapter 7	Spherical avatar telepresence system	Avatar	Trust for a single observer at multiple viewpoints

Chapter 4

Experiment: Gaze in spherical video telepresence system

The overarching goal of the three evaluations on the affordance of object focused gaze in telepresence systems documented over the following three chapters is to investigate whether our displays can more faithfully represent the gaze of the remote user, comparing to previous telepresence systems (see chapter 2). The current chapter investigates a single observer at multiple viewpoints, Chapter 5 examines multiple observers at multiple viewpoints, and Chapter 6 explores the effect of adding stereoscopy for multiple observers at multiple viewpoints.

The two experiments presented in this chapter evaluated the spherical video telepresence system. We compared the effectiveness of both spherical and flat displays by measuring the ability of observers to accurately judge which target a user is gazing at. Experiment 1, a pilot study, demonstrated the spherical video telepresence system can convey gaze relatively accurately. Experiment 2 compared observers' performance in different flat and spherical display conditions by modeling systematic biases and investigating the influence of seat and target positions.

4.1 Experimental design

In order to evaluate teleconferencing systems, several independent variables, explained in the Table 4.1, should be taken into consideration in experiment design. In particular, we explore the influence of display modes, seat positions, target positions. The setup of our experiments and independent variables are presented below.

Table 4.1: Factors for evaluating teleconferencing systems

Technical quality	Latency during the video transmission Sensitivity of eye tracking
	Number of capturing cameras to present continuous head rotation of remote user (view interpolation)
	Size of video Image
	Cost of system
	Lighting
	Position of capturing camera
Independent variable	Viewers' positions relative to sphere display, namely, angle and distance
	Number of viewers
	Remote user's position relative to capturing cameras
Conversation quality	Mutual eye gaze
	Gaze direction to indicate whom they are addressed or suggested
	Gaze direction to indicate which object they are pointed out
	Social engagement, such as, telling truth/ lie
	Learning effects

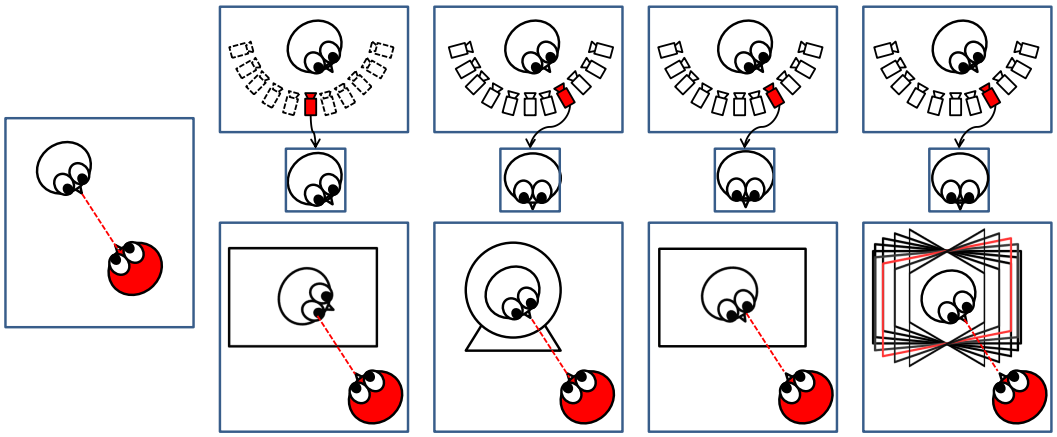
Figure 4.1: Capture system: The actor gazes at the target card 13 captured by semicircular camera arrays in remote room.

Figure 4.2: Display system: Since the principal observer is seating in viewpoint N=9, the video captured by camera N=9 is presented on the sphere display, which lines up with the principal observer N=9. (Also see Figure 3.4)

4.1.1 Setup

We used a half annulus table with the larger semicircle of radius 1405mm and the smaller one of radius 405mm (see Figure 4.1 and Figure 4.2). Horizontally, 23 gaze target cards were placed in a semicircle of radius 905mm, every 7.5° on the table. When capturing video, the target cards with even numbers lined up with the cameras and the observer's viewpoints.

4.1.2 Independent variable



(a) Face to face (b) Fixed single video flat display (c) Sphere display (d) Fixed multiple video flat display (e) Free multiple video flat display

Figure 4.3: Five levels of categorical variable media representation. The observer (in red) is seated at viewpoint 4, therefore camera 4 (in red) is enabled. Top row: capturing actor in the remote room; middle row: captured video for transmission; bottom row: view of screen showing actor's gaze direction in the local room. The dashed red line is the actual actor's gaze direction

4.1.2.1 Display modes

The display mode variable consists of five display types: Face to face (*Face*), sphere display (*Sphere*), fixed single video flat display (*Fixed single flat*), fixed multiple video flat display (*Fixed multiple flat*) and free multiple video flat display (*Free multiple flat*). We ensured that the vertical alignment of the eye gaze of the actor, the eye level of observers, eye level of the video of the actor on the spherical or fixed single video flat display, capturing cameras, and attention target cards were the same. This ensured equivalence in stimuli alignment and apparent size between the four display conditions and the face to face condition.

Note that although the system as designed and built is a real-time collaborative system that can connect a remote room to a local room, video was recorded to disk and replayed for the purposes of control of the experimental stimuli in these four display conditions, with exception of face to face condition.

- Face to face

  Figure 4.3(a) shows the face to face condition, where the observer and actor were in the same room. The actor sat at the center position of the table and the observer sat on the outside. The actor was wearing small headphones listening to the same audio instruction as was used when recording the videos for the display conditions.

- Fixed single video flat display

  The spatial arrangement of this condition was identical to the sphere display condition except the conventional flat display and only the center camera (lined up with position 6) was used, depicted in Figure 4.3(b). Image quality remained the same. This condition mimicked the commonly found distorted video conferencing system where the actor is not always lined up with the capturing camera, and the observer is not always lined up with the display screen.

- Sphere display

  In Figure 4.3(c), the observer observed the pre-recorded video on our spherical video telepresence system (see Section 3.1). Hence the actor and the observer achieved line of sight effect.

- Fixed multiple video flat display

  This condition was similar to the fixed single video flat display condition except all the capturing cameras were used, presented in Figure 4.3(d). According to the observer's position, the proper video is selected. The actor is always lined up with the capturing camera, but the observer might be looking obliquely at the screen.

- Free multiple video flat display

  This condition was similar to the fixed multiple video flat display condition except the flat display is rotated based on the observer's position allowing observer directly looking at the screen, shown in Figure 4.3(e). Hence the actor and the observer achieved the line of sight effect.

4.1.2.2 Seating positions

We define the participants' seating positions at 30°, 45°, 60°, 75°, 90°, 105°, 120°, 135° and 150° relative to the display. Therefore there were 9 levels of categorical variables of seat position. The distance between participant and display remained constant.

4.1.2.3 Target numbers

Twenty three numbered target cards were placed on the semicircular table from 15° to 165° at every 7.5°. Therefore there were 23 levels of categorical variable of target numbers. The distance from target position to participant and display remained constant.

4.2 Experiment 1

The purpose of the first experiment was to demonstrate that the combination of a spherical display and a camera array can better represent the actor's gaze than a fixed single video flat display. We measured the effectiveness of the displays by measuring the ability of observers to accurately judge which target the actor was gazing at for three display modes, presented in Figure 4.3(a), 4.3(b) and 4.3(c). Also, we investigated the situation that the observer was not seated in the same direction as the camera that was observing the actor. We formed two hypotheses.

4.2.1 Hypothesis

4.2.1.1 Hypothesis 1

It is hypothesized that both face to face and sphere display will demonstrate higher levels of accuracy (the observer is accurate if they successfully identify the correct target) than fixed single video flat display when the observers are in varied positions. We further expect face to face to be better than sphere display.

4.2.1.2 Hypothesis 2

For both sphere display and fixed single video flat display, it is hypothesized that if the observer is not seated in the same direction as the camera that is observing the actor, the accuracy will be worse than if the camera chosen for the display is aligned with the observer's position.

4.2.2 Method

4.2.2.1 Participants

60 participants, students and staff at University College London, were recruited to take part as observers in this user study. 20 groups of three were used for testing and each group experienced one of three different conditions (sphere display, fixed single video flat display, face to face). Eight further confederates were actors in these experiments: four actors were recorded on video for the sphere and fixed single video flat condition and four acted in the face to face condition.

4.2.2.2 Apparatus and materials

For the two display conditions, we video-recorded the actors' head movements, presented in Figure 4.1. The actor sits at the center position of the half annulus table and his or her head is captured by 11 video cameras. The actor listens to an audio recording that instructs them to look at the gaze target cards. A new target is given every 10 seconds. The targets are randomly ordered, and each one is gazed at twice, amounting to 46 targets in the audio instruction and thus in the recorded videos. Four participants were actors, and thus four sets of 11 videos were generated.

4.2.2.3 Procedure

Nine different positions for observers were investigated. Observers took part in groups of three. In all conditions, the group performed three trials. On each trial, the group would sit in positions 2,3 & 4 or 5,6 & 7 or 8, 9 & 10.

For each trial, each observer was given a sheet of paper with an empty grid of 46 squares. In all three conditions, the actor or the video of the actor reoriented to a new target card every 10 seconds. At the same time an audio prompt to the observers instructed them that this was a new target. They would then judge which target (1-23) the actor was gazing at and then write this in the relevant grid square.

For the face to face condition, the three observers and actor were in the same room. The actor sat at the center position of table and the three observers sat on the outside. The actor was wearing small headphones listening to the same audio instruction as was used when recording the videos for the display conditions. The actor performed the sequence of gazes three times. On each repetition, the group of three observers moved to another one of the group positions.

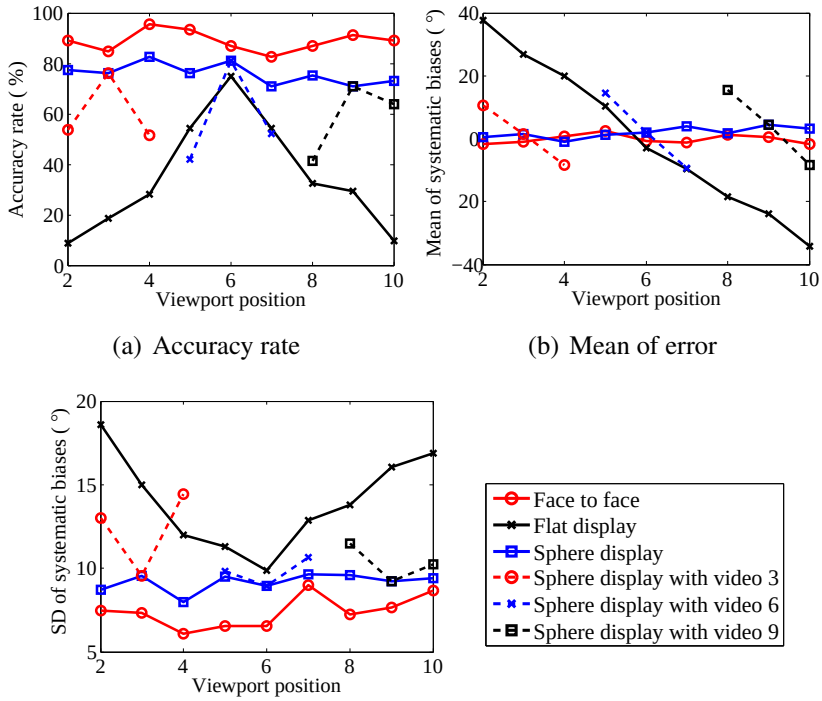
For the sphere display condition, the three observers observed the pre-recorded video on the sphere display, presented in Figure 4.2. For each group position, one of the observers was the principal observer. The video corresponding to the principal observer's position was shown on the display. Each group saw the actor's video three times. On each repetition, the group of three observers exchange positions, hence each observer became a principal observer at least once.

For the fixed single video flat display condition, the three observers observed the pre-recorded video on the fixed single video flat display. The video was always from camera position six, simulating a simple web-cam set up where the observers might be looking obliquely at the screen, and the actor looking obliquely at the camera.

The experiment took about 20 minutes.

4.2.3 Results

A summary of the results of the experiment are presented in Figure 4.4. In each figure, the horizontal axis indicates the viewpoint position (p) from 2 to 10. The angle of viewpoint position(α) in degrees is from 30° to 150° at every 15° relative to center of



(a) Accuracy rate

(b) Mean of error

(c) Standard deviation of error

Figure 4.4: Result for analysing the actual targets and perceived targets in different treatment conditions.

the conferencing table.

$$\alpha = p \times 15^{\circ}. \tag{4.1}$$

The primary measurement in our results is the accuracy rate in perceiving the attention target. The accuracy rate is percentage of accurate prediction over total prediction.

We then define systematic bias (β_i) to be the difference between the actual target number (t_{ai}) and the observer's perceived attention target number (t_{oi}) converted to degrees, based on attention targets being 7.5° apart from each other.

$$\beta_i = (t_{ai} - t_{oi}) \times 7.5^{\circ}. \tag{4.2}$$

Each observer indicates 46 target positions in each trial. Each observer does three trials. There are 12 observers in the face to face condition (four groups of three) and nine observer seat positions. Thus, there are 184 ($46 \times 3 \times 12/9$) rating events in each seating position. Similarly, there are 184 rating events in each seating position for the

fixed single video flat display. For the sphere display, there are 36 observers (twelve groups of three) but only one of the group is in the principal position. Thus, there are also 184($46 \times 3 \times (36/3)/9$) rating events for principal observers in each of the nine observer seating positions. However in the analysis below, we include some data from the secondary observers. In particular, for seating positions 3, 6 and 9, we analyze the 184 rating events for the observer seated on their left and 184 rating events for the observer on their right. This gives us a view of how important it is to use the correct video for the observer position.

4.2.3.1 Accuracy rate

The result of accuracy rate in different conditions is shown in Figure 4.4(a). For the fixed single video flat display, with the observer at the central viewpoint, the accuracy rate is 75%. However, the accuracy rate drops off symmetrically as the observer position diverges from the central position. This is expected as when the observer is not seated in position 6, they still see the video taken from the camera at position 6.

The results for face to face and sphere display are not affected by viewpoint position and the average accuracy rates are 89% and 76%, respectively. The average accuracy rate of sphere display is slightly lower than face to face, but similar to the observer sitting at the central position in the fixed single video flat display condition. The fact that the accuracy does not vary with observer position for the sphere display when considering the principal observer supports the primary hypothesis. The performance of the sphere display at the extreme positions (2 and 10) is significantly above that of the fixed single video flat display.

When we consider the secondary positions in the sphere display, the three "three point hat" graphs in Figure 4.4(a), we see that it is very important that the camera selected be aligned with the observer position. Considering the principal observer at position 3, we see that the observer in position 2, observing the video from position 3, has a performance of under 54% compared to the accuracy of almost 76% for the principal observer seated immediately to their right. This pattern is repeated for all secondary observers.

The difference between face to face performance and sphere display performance may be due to video quality. We note that for observer position 6 on the fixed single

video flat display, the ideal situation for this position, the accuracy is very similar to the sphere display at this position. This indicates that the sphere display is no worse than the fixed single video flat display, but it has the advantage that it has the same apparent size in the different observer positions.

4.2.3.2 Mean of systematic biases and standard deviation

Next, we analyzed the mean and standard deviation of systematic biases for the actual targets and observers' perceived targets in different treatment conditions, shown in Figure 4.4(b) and Figure 4.4(c). For the face to face and sphere display conditions, the observer position has no significant effect on the mean of systematic biases which is around 0°. The standard deviation of the systematic biases for the sphere display is higher, but there are no systematic biases, indicating that the observers are generally finding it harder to determine gaze.

In contrast, for the fixed single video flat display, the mean of systematic biases varies linearly according to viewpoint position. We utilized the first-order Matlab® Polyfit function to generate the coefficients of the polynomial to simulate a curve to fit the data and found a relationship between the systematic biases of mean and angle of viewpoint position:

$$\sigma(\beta_i) = -0.6\alpha + 54.27^\circ = 0.6 \times (90^\circ - \alpha) + 0.27^\circ. \tag{4.3}$$

The linear model of systematic biases in the fixed single video flat display condition is interesting in that it suggests that the observer's judgment of gaze angle from front is only 60% of what it should be. Therefore, for the fixed single video flat display, the observer perceives the actor to be looking more directly straight out of the display.

4.3 Experiment 2

In the second experiment, we introduced two more display modes, shown in Figure 4.3(d) and 4.3(e). We compare the sphere display with fixed multiple video flat display and free multiple video flat display to demonstrate the improvement of representing the actor's gaze by using the camera array and the spherical display simultaneously. In addition, we used the mixed design Analysis of Variance (ANOVA) as a more reliable statistical analysis to further investigate factors influencing the observers

in perceiving targets in different conditions. We specifically formed three hypotheses.

4.3.1 Hypothesis

4.3.1.1 Hypothesis 3

We explored the level of error with which observers can discriminate the actor's gaze orientation for all five display modes. Specifically we measured the ability of participants to identify which set of targets the actor appears to be gazing towards. Given the five display modes, we expected that the level of error of observers' performance would follow the trend below:

$$Face < Sphere < Free\ multiple\ flat < Fixed\ multiple\ flat < Fixed\ single\ flat. \tag{4.4}$$

4.3.1.2 Hypothesis 4

We then explored the influence of seat position. We expected that face to face, sphere display and free multiple video flat displays will show a similar level of error for all seat positions. However, the level of error will increase symmetrically as the observer position diverges from the central position for fixed multiple video flat display and fixed single video flat display.

4.3.1.3 Hypothesis 5

We further explored the influence factor of target position. We expected that face to face, sphere display and free multiple video flat displays will show similar level of error while observing all numbered targets. However, there should be systematic biases for fixed multiple video flat display and fixed single video flat display.

4.3.2 Method

4.3.2.1 Participants

40 participants, students and staff at University College London, were recruited to take part as observers in our user study. Each participant judged only one of five display modes, a between-subjects design. However, a within-subjects design was employed for the two factors of 9 seating position (2-10) and the 23 target numbers (1-23). We randomly mixed the seating positions and target numbers in order to reduce any confounding influence of the orderings such as learning effects or fatigue.

Two further participants were actors in this experiment: one actor was recorded on video for four video display conditions and the other acted in the face to face condition.

4.3.2.2 Apparatus and materials

For the four display conditions we recorded the actor's performance. The actor sits at the center position of the half annular table and his or her head is captured by 11 video cameras. The actor listens to an audio recording that instructs them to look at the gaze target cards. A new target is given every 10 seconds. The targets are randomly ordered, giving 23 targets in the audio instruction and thus in the recorded videos. A set of eleven videos were generated.

4.3.2.3 Procedure

The experiment took about 30 minutes for each participant. Upon arrival, each participant was assigned to one of five treatment conditions. Eight observers are investigated for each treatment condition.

Nine different positions for each observer were investigated. Observers were initially seated in one of the nine positions in a counterbalanced random order. For each trial, each observer was given a sheet of paper with an empty grid with 23 squares. Every 10 seconds, the actor reoriented to a new target card. At the same time, an audio prompt to the observers instructed them that this was a new target. They would then judge which target (1-23) the actor was gazing at and write this in the relevant grid square. After each trial, the session was paused to allow the participants to change seating position accordingly.

4.3.3 Results

4.3.3.1 Level of error

The primary measurement in our results is the level of error in perceiving the attention target. We define error (ε_i) to be the absolute value of difference between the actual target number (t_{ai}) and the observer's perceived attention target number (t_{oi}) converted to degrees, based on attention targets being 7.5 $^\circ$ apart from each other.

$$\varepsilon_i = |t_{ai} - t_{oi}| \times 7.5^\circ. \tag{4.5}$$

The dependent variable data (ε_i) were entered into a mixed design Analysis of

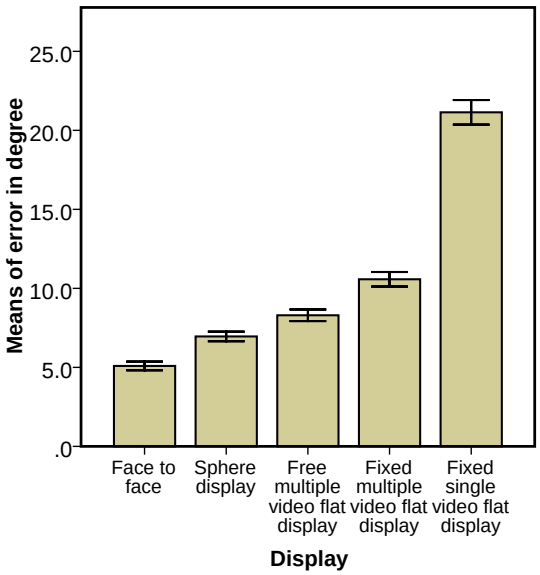


Figure 4.5: Bars show estimated marginal means of error in different treatment conditions, error bars show 95% CI of the means

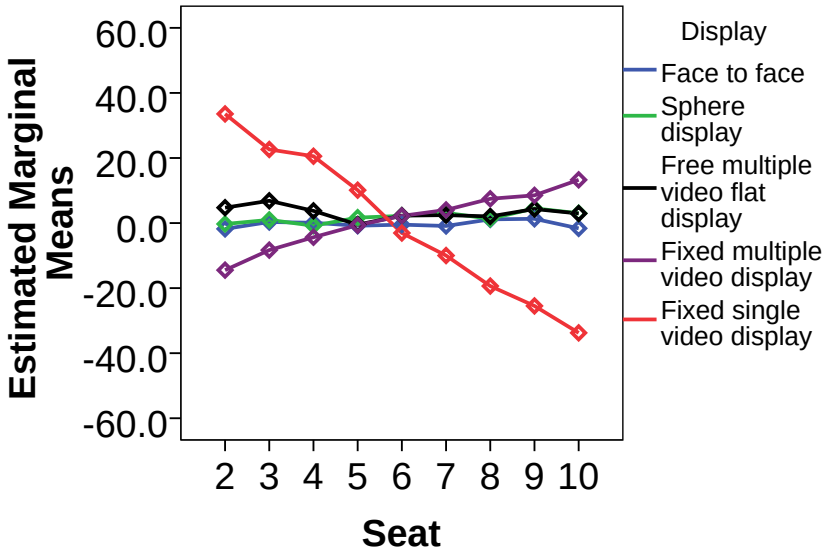


Figure 4.6: 2-way interaction: estimated marginal means of biases in degree

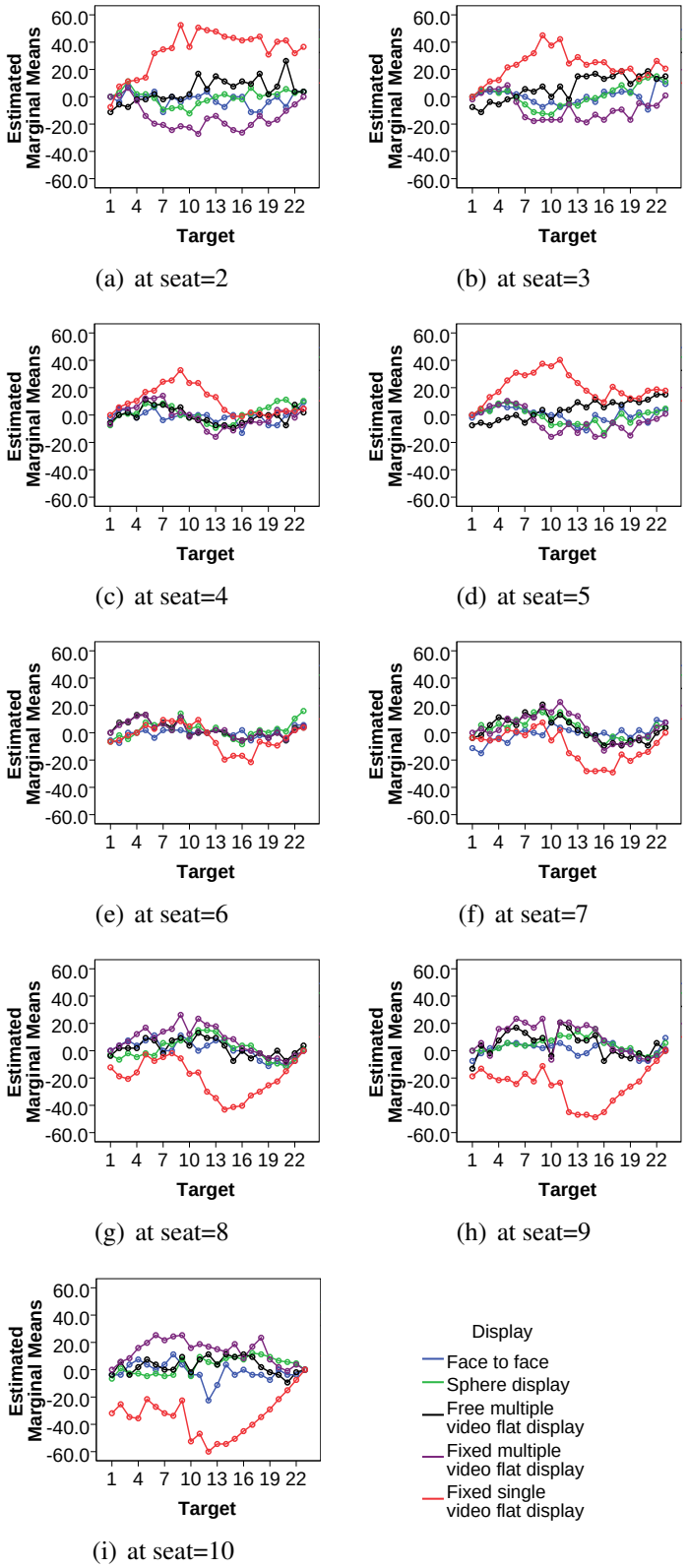


Figure 4.7: 3-way interaction: estimated marginal means of biases in degree

Variance (ANOVA) with the three factors of display condition, seating position, and target position. We used Mauchly's test of sphericity to validate our repeated measures factor ANOVAs, thus ensuring that variances for each set of difference scores were equal. Mauchly's test indicated that the assumption of sphericity had not been violated.

Results reveal that there was a significant main effect of display condition, $F(4,8279) = 684.842, p < 0.01$ and Post-hoc Tukey tests revealed significant mean differences between each of all those displays. The face to face ($Mean$, $M = 5.104$) achieved the lowest level of error, followed by sphere display ($M = 6.916$), free multiple video flat display ($M = 8.262$), fixed multiple video flat display ($M = 10.375$), and then fixed single video flat display ($M = 21.162$). See Figure 4.5. This supports the third hypothesis.

While this absolute level of error is a good basic measure, it effectively accumulates the positive and negative systematic biases. In order to get a more detailed view of effectiveness of different display in perceiving the attention target, whether there is left or right systematic biases, how seat position varies and target position variable effect, the result of different display condition must be taken into account.

4.3.3.2 Systematic biases

Similarly, we then looked into systematic bias (β_i), which is defined in the first experiment. Firstly, we look into 2-way interaction. Figure 4.6 shows the average systematic bias of different seat positions under five different display conditions. For face to face, sphere display and free multiple videos flat display, the average systematic bias curves roughly around 0 degree and did not change over different seating positions. Moreover, the face to face condition is the most stable and the closest approximate to 0 degree, followed by sphere display and then the free multiple video flat display. By contrast, the average systematic bias varies linearly according to seat position for fixed multiple video display and fixed single display. The absolute value of systemic bias is the error which is defined above. The lines of fixed multiple video flat display and fixed single video flat display are symmetric about $seat = 6$. Therefore, the seat variable only has an effect for fixed multiple video condition and fixed single video condition. This supports the fourth hypothesis.

We conducted a 3-way interaction to investigate whether the seat $\times$ display in-

teraction described above is the same for all targets. We used the estimated marginal means to interpret the 3-way interaction (Figure 4.7). For face to face, sphere display and free multiple videos flat display, the average systematic bias curves are basically around 0 degree with slight fluctuations among different target positions.

However, for the fixed multiple videos flat display and the fixed single flat display, the average systematic bias varies over different target positions. The fixed single video display has more biases compared to the fixed multiple videos display. This supports the fifth hypothesis.

Interestingly, Figure 4.7 shows that the curves can be modified into symmetrical parts for each pair of seat positions 2 & 10, 3 & 9, 4 & 8 and 5 & 7, which are symmetrically arranged on both sides of the center seat position 6. For seat position 6, the curve itself is symmetry relative to point $(12,0)$.

4.3.3.3 Linear regression for systematic biases

As discussed in the previous section, the mean of systematic biases varied linearly according to viewpoint position for the fixed single video flat display (red line in Figure 4.6) and multiple video flat display (purple line in Figure 4.6) conditions. However, the mean of systematic biases are sloped in opposite directions in those two conditions.

A simple regression was carried out to ascertain if the angle of viewpoint position (α) can predict the systematic biases of fixed single video flat display $(\beta_{fixed\ single\ flat})$. A strong correlation was found between the angle of viewpoint position and the systematic biases of fixed single video flat display, $r = .831$ and the regression model predicted 69% of the variance. The model was a good fit for the data, $F(1,1654) = 3685.526, p < .001$. The linear regression model is presented in Equation 4.6, $b = -.57$, $t(1654) = -60.709, p < .001$. This further confirmed the result in Equation 4.3 in the first experiment.

Similarly, standard simple regression analysis was conducted to evaluate how well the angle of viewpoint position (α) predicted the systematic biases of fixed multiple video flat display $(\beta_{fixed\ multiple\ flat})$. The angle of viewpoint position was significantly related to the systematic biases of fixed multiple video flat display, $F(1,1654) = 814.257, p < .001$. The correlation coefficient was $r = .574$, indicating that approximately 33% of the variance of the systematic biases of fixed multiple video flat dis-

play can be accounted for by angle of viewpoint position. The regression equation for predicting the systematic biases of fixed multiple video flat display was shown in Equation 4.7 , $b = .221,\ t(1654) = 28.535, p < .001$.

$$\beta_{fixed\ single\ flat}(\alpha) = -0.57\alpha + 50.804^\circ = 0.57 \times (90^\circ - \alpha) - 0.496^\circ. \quad (4.6)$$

$$\beta_{fixed\ multipl\ flat}(\alpha) = 0.211\alpha - 18.13^\circ = -0.211 \times (90^\circ - \alpha) + 0.86^\circ. \quad (4.7)$$

4.4 Discussion

4.4.1 Camera arrays vs. single camera

The line of fixed single video flat display has a higher slope value compared to fixed multiple video flat display (in Figure 4.6). This indicates a steeper incline and higher systematic biases. In some extreme cases, such as, in seat positions 2 and 10, the observer had more difficulty in perceiving targets in fixed single video flat display. The fixed multiple video display improves the system's ability to represent the actor's gaze, by lining up the capturing cameras using camera arrays.

4.4.2 Directional projection

The gradient of line indicates systematic biases in fixed multiple video flat display (Figure 4.6) however, the line is always stable around 0 degree for the free multiple video flat display. The observer can perceive targets better in free multiple video flat display, particularly, when seat position is further apart from the center. The free multiple video flat display improves the system's ability to present the actor's gaze, by providing perspective correct projection.

4.4.3 Sphere vs. free multiple video flat display

Figure 4.5 shows that the level of error in the sphere display is only slightly lower than the free multiple video flat display condition. However, in free multiple video flat display, we have to manually rotate the flat display for each viewpoint position for each observer, which is impossible for practical video conferencing.

Previous findings [4, 81] suggested that biases occur differently while observing convex, flat and concave surfaces. For this spherical display, we plan to further explore this finding, with our next step being to collect data for more viewing angles.

4.4.4 Video quality

The higher level of error in Figure 4.5 and larger fluctuation around 0 degree in Figure 4.6 in sphere display compared to face to face shows that observer can better perceive the actor's attention target in face to face. This suggests that there is more work to be done on the quality of representation of gaze with such displays.

4.4.5 Seat position

From the discussion above, the seat position has a linear effect on the fixed single flat display and fixed multiple video display. Observers could interpret the direction of actor gaze of the sphere display more accurately than the free multiple video flat display and similarly to the face to face condition for all seat positions.

4.4.6 Linear model for predicting distortion

The study by Roberts et al [94] found that the correct viewing of the sides of the face is important for the interpretation of gaze. Large errors in estimation coincided with either the face being viewed from the wrong perspective or unevenly lit. This is inline with our results, from which we modeled the systematic biases for two flat display configurations. We found the negative linear correlation between the angle of viewpoint position and the systematic biases of the fixed single video flat display in Equation 4.6, and the positive linear correlation between the angle of viewpoint position and the systematic biases of the fixed multiple video flat display in Equation 4.7, respectively. This indicates that the fixed single video flat display is biased in the opposite direction to the fixed multiple video flat display condition (see Figure 4.3). Whilst the biases may have been caused by incorrect viewing angles in both conditions, the single capturing angle of the fixed single video condition may have caused the bias to be in the opposite direction. Also, this effect appears very reliable and this means that it may be possible to model and thus predict the distortion.

4.5 Chapter summary

The two experiments presented in this chapter evaluated a spherical video telepresence system by measuring the ability of observers to accurately judge which targets the actor is gazing at. Results from the first experiment demonstrate the effectiveness of the camera array and spherical display system, in that it allows observers at multiple observing positions to accurately tell which targets the remote user is looking at. The second experiment further compared a spherical display with a planar display and provided detailed reasons for the improvement of our system in conveying gaze. We found two linear models for predicting the distortion introduced by misalignment of capturing cameras' and observer's viewing angles in video conferencing systems.

Chapter 5

Experiment: Gaze in cylindrical video telepresence system

This chapter presents an experiment to test if the cylinder multiview system (see Section 3.3) can better represent the remote person's gaze for multiple observers. We measured the effectiveness of the displays by measuring the ability of multiple observers to accurately judge which target the remote person was gazing at.

5.1 Experimental Design

5.1.1 Display conditions

We compared four display conditions. *Cylinder multiview multi-video* condition was our system discussed in Section 3.3, which could support correct viewing for multiple viewpoints around a conference table (see Figure 5.1(a)). *Cylinder multiview single-video* condition was identical to the cylinder multiview multi-video condition, except that only the center camera was used for capturing the remote person. All projectors projected this video, instead of projecting unique perspective-correct videos. Thus, observers would perceive the gaze direction as if they were standing straight in front (see Figure 5.1(b)). *Cylinder diffuse single-video* condition used a curved diffuse white projection screen. Only the center camera and projector were used (see Figure 5.1(c)). *Flat diffuse single-video* condition used a conventional 2D flat screen, instead of 3D cylinder surface. This condition mimicked the commonly found the Mona Lisa gaze effect, which occurs when 3D objects are rendered in 2D, causing the gaze perception of all in a room to be the same (see Figure 5.1(d)). Image quality remained the same in

(a) Cylinder multiview multi-video condition (b) Cylinder multiview single-video condition

(c) Cylinder diffuse single-video condition (d) Flat diffuse single-video condition

Figure 5.1: Photos of display conditions taken from viewpoint 1: when the remote person gazing at the target 10, observers perceive different targets in four display conditions.

all conditions.

5.1.2 Viewpoints

We explored four observers' viewpoints (1, 4, 5 & 9). We included viewpoint 5 where the observer at the center position as a benchmark; viewpoint 1 and 9 where observers sat at two extreme viewing angles; and viewpoint 4 where the observer sat right next to observer 5.

5.2 Experiment

5.2.1 Hypothesis

5.2.1.1 Hypothesis 1

We expected a similar level of error for observer perceiving targets at all viewpoints in the cylinder multiview multi-video condition. we expected the level of error will increase symmetrically as the viewpoint diverges horizontally from the central position for the other three display conditions.

5.2.1.2 Hypothesis 2

We expected that observers in cylinder multiview single-video condition and flat diffuse single-video condition will identify much more incorrect targets compared to those in cylinder multiview multi-video condition. We further expected the cylinder diffuse single condition to lie between these two in performance, as the 3D cylindrical surface eliminates the Mona Lisa effect [4] but observers could only see part of head in some extreme viewpoints.

5.2.2 Method

5.2.2.1 Participants

48 participants, students and staff at University College London, were recruited to take part as observers in our user study. All participants had normal or corrected to normal eye sight. One further participant was a remote person recorded on video.

The experiment had a 4 display conditions$\times$4 viewpoints$\times$15 target positions mixed design, with a within-subjects design for target positions but a between-subject design regarding display modes and viewpoints.

5.2.2.2 Apparatus and materials

We video-recorded the remote persons' head movements (see Figure 3.15(a)). The remote person sat at the center position of the table and his or her head is captured by 4 cameras simultaneously. The remote person listened to an audio recording that instructed to turn his or her head to look at the targets. A new target was given every 10 seconds. The targets were randomly ordered, each one was gazed at only once, amounting to 15 targets in the audio instruction and thus in the recorded videos. One

set of 4 videos were generated.

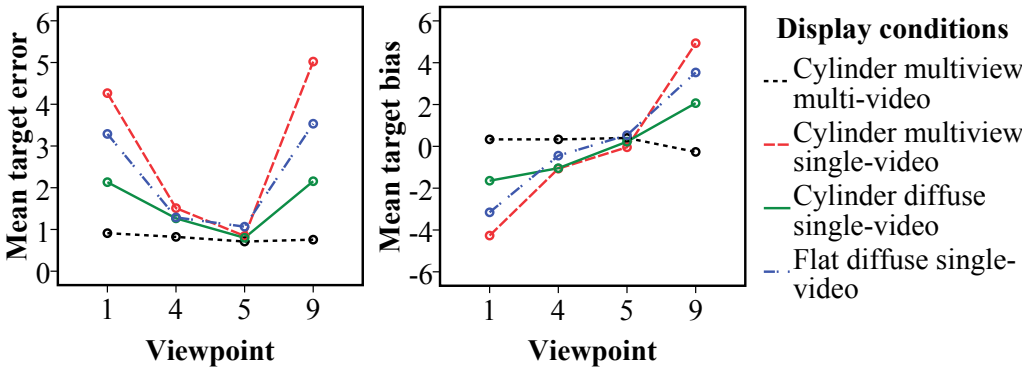
5.2.2.3 Procedure

12 groups of four were used for testing, and each group experienced one of four different display conditions with each observer sat at one of the four viewpoints (see Figure 3.15(c)). Each observer was given a sheet of paper with an empty grid of 15 squares. The video of the remote person reoriented to a new target card every 10 seconds. At the same time an audio prompt to the observers instructed them that this was a new target. Then, observers would judge which target (1-15) the remote person was gazing at and then write this in the relevant grid square. The experiment took about 5 minutes. Participants received chocolates as compensation.

5.2.3 Result

The primary measurement in our results was the level of error in perceiving targets. We defined target error (ε_i) to be the absolute value of difference between the observer's perceived target number (t_{oi})and the actual target number (t_{ai}): $\varepsilon_i = |t_{oi} - t_{ai}|$. Figure 5.2(a) shows the target error at the four viewpoints in four display conditions. The line of the cylinder multiview multi-video condition shows that it achieved the lowest mean target error. The means were very similar across the four viewpoints, indicating that the viewpoint had little impact in this display conditions. At the extreme viewpoints (1 and 9), the means were significantly below that of the other three display conditions. In addition, the graph shows that the central viewpoint had the lowest mean target error, where four display conditions all had perspective-correct video; the mean target error increased symmetrically as the viewpoint diverges from the central position for cylinder multiview single-video condition, cylinder diffuse single-video condition and flat diffuse single-video condition. This is expected as when the observer did not sit in viewpoint 5, those display conditions still used the video from camera 5.

A 4 display conditions $\times$ 4 viewpoints $\times$ 15 target positions mixed design ANOVA was conducted on the target error, with display condition and viewpoints as two between-subjects factors and target positions as a within-subjects factor. Mean of target error differed significantly across the four display conditions, $F(3,32) = 32.167, p < .001$. Tukey post-hoc tests revealed significant mean differences between each of the display conditions. The cylinder multiview multi-video condition ($M =$



(a) Mean target error (b) Mean target bias

Figure 5.2: The mean target error and mean target bias for each display conditions and viewpoints.

$.800, 95\%\ CI\ [.473, 1.127]$) gave significantly lower mean target error than the cylinder diffuse single-video condition ($M = 1.589, 95\%\ CI\ [1.262, 1.916]$), $p = .008$, the cylinder multiview single-video condition ($M = 2.911, 95\%\ CI\ [2.584, 3.238]$), $p < .001$, and the flat diffuse single-video condition ($M = 2.294, 95\%\ CI\ [1.968, 2.621]$), $p < .001$. This supports the primary hypothesis. Results also revealed a significant main effect of viewpoints, $F(3, 32) = 39.448, p < .001$. Tukey post-hoc comparisons indicated the mean target error at viewpoint 5 ($M = .856, 95\%\ CI\ [.529, 1.182]$) is significantly lower than viewpoint 1 ($M = 2.65, 95\%\ CI\ [2.323, 2.977]$), $p < .001$ and viewpoint 9 ($M = 2.867, 95\%\ CI\ [2.54, 3.194]$), $p < .001$, which supports the second hypothesis; however, it did not significantly differ from viewpoint 4 ($M = 1.222, 95\%\ CI\ [.895, 1.549]$), $p > .05$, which is expected as the seat position only slightly diverges from the front. The mean target error at viewpoint 1 did not significantly differ from viewpoint 9, $p > .05$, which is also expected as the viewing angles of viewpoint 1 and 9 are equal only opposite in direction. The display conditions $\times$ viewpoints interaction was significant, $F(9, 32) = 7.277, p < .001$, indicating that mean of target error due to viewpoints were different in four display conditions.

We further investigated whether there was leftward bias or rightward bias in perceiving targets in different display conditions. We defined target bias (β_i) to be the difference between the observer's perceived target number (t_{oi}) and the actual target number (t_{ai}): $\beta_i = t_{oi} - t_{ai}$. Figure 5.2(b) shows the target error at four viewpoints in four display conditions. Positive values indicated leftward biases; whereas negative

values indicated rightward bias. For the cylinder multiview multi-video condition, the mean target bias did not change substantially across different viewpoints. This further supports the hypothesis. By contrast, for the other three display conditions, the biases were dependent on observers' viewpoints. For the flat diffuse single-video condition, the biases of four viewpoint in this study nicely fit in the previous work [80] that is the mean target bias varies linearly according to seat position. The graph also shows that the bias of cylinder diffuse single-video condition is less than flat diffuse single-video condition. This parallels the previous finding [4] that biases occur differently while observing convex, flat and concave surfaces.

5.3 Chapter summary

The experiment reported in this chapter evaluated the effectiveness of our cylindrical video telepresence system by measuring the ability of observers to accurately judge which target the remote person is gazing at. We compared our system to three alternative display configurations. We ran an experiment to demonstrate that our system can convey gaze relatively accurately, especially for observers viewing from off-center angles. This demonstration and results thus motivate the further study of novel display configurations and the supporting camera and networking infrastructure for them.

Chapter 6

Experiment: Head gaze in random hole autostereoscopic multiview telepresence system

In this chapter, we investigated using the random hole display to represent remote person. The gaze direction can be influenced by many visual components, such as, head orientation and orientation of the eyes relative to the head. This study explores the effectiveness with which observers can discriminate an avatar's head orientation when the avatar's eyes are centered in the head, because head gaze is a good indicator of focus of attention in human computer interaction applications. We evaluated this system by measuring the ability of observers with different horizontal and vertical viewing angles to accurately judge which targets the avatar is gazing at. We compared 3 perspective conditions: a conventional 2D view, a monoscopic view with motion parallax, and a stereoscopic view with motion parallax. Although the random hole display does not provide high quality view comparing to other display technologies, the unique view content is easily distinguished. Results suggest that the combined presence of motion parallax and stereoscopic cues significantly improved the effectiveness with which observers were able to assess the avatars gaze direction. This motivates the need for stereo in future multiview displays.

6.1 Experimental Design

The purpose of the experiment was to demonstrate that the random hole telepresence system can better represent the remote person's gaze for multiple observers. We mea-

sured the effectiveness of the display by measuring the ability of multiple observers to accurately judge which target the avatar was gazing at.

We compared 3 perspective conditions. For the conventional 2D condition, the conventional display was shown from the perspective of a front facing camera, centered on the avatar's head. This condition mimicked the commonly found Mona Lisa gaze effect. For the motion parallax condition, the random hole display was displayed with perspective correct monoscopic view based on the location of the observer relative to the display. For the motion parallax & stereoscopy condition, the random hole display was displayed with correct perspective for each of observers' eyes, that provided them with a fully stereoscopic image, giving the impression that the avatar's head was inside the display. The apparent size of avatar remained the same in all conditions.

We explored 9 observers' viewing angles, including three horizontal viewing angles ($-30°$, $0°$ & $+45°$) and three vertical viewing angles ($-10°$, $0°$ & $+20°$). The two extreme vertical viewing positions are where the observer sat right on the floor ($-10°$) and the observer stood up straight ($20°$).

6.2 Experiment

6.2.1 Hypotheses

6.2.1.1 Hypothesis 1a

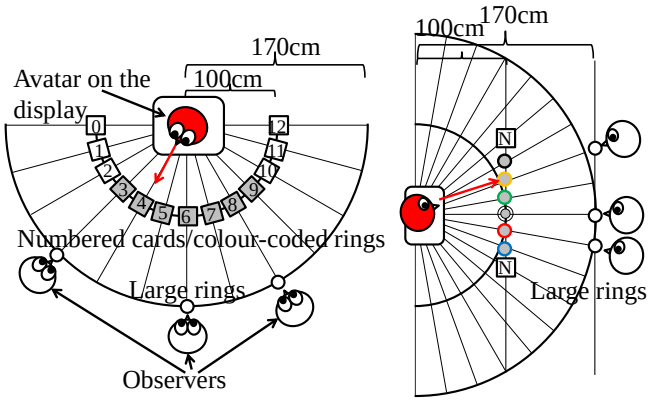
Horizontally, we expected that the participants will introduce the lowest level of error when identifying correct targets in the motion parallax & stereoscopy condition, followed by the motion parallax condition and then the conventional 2D condition.

6.2.1.2 Hypothesis 1b

Vertically, we expected that the participants will introduce the lowest level of error when identifying correct targets in the motion parallax & stereoscopy condition, followed by the motion parallax condition and then the conventional 2D condition.

6.2.1.3 Hypothesis 2a

Horizontally, we expected the level of error for observer perceiving targets at all horizontal viewing angles remain stable in both the motion parallax & stereoscopy condition and the motion parallax condition. However, the level of error will increase as the



(a) Horizontal view (b) Vertical view

Figure 6.1: Schematic layout of experiment setup. Note that the gray area covered actual target positions.

	$-45°$	$-30°$	$-15°$	$0°$	$+15°$	$+30°$	$+45°$
$+20°$	18	21	13	22	19	32	33
$+10°$	1	16	14	25	6	17	20
$0°$	15	3	7	4	35	23	27
$-10°$	9	26	34	29	31	10	12
$-20°$	2	11	28	30	24	5	8

Table 6.1: Target Order

viewing angle diverges horizontally from the central viewing angle for the conventional 2D condition.

6.2.1.4 Hypothesis 2b

Vertically, we expected the level of error for observer perceiving targets at all vertical viewing angles remain stable in both the motion parallax & stereoscopy condition and the motion parallax condition. However, the level of error will increase as the viewing angle diverges vertically from the central viewing angle for the conventional 2D condition.

6.2.2 Method

6.2.2.1 Participants

27 participants, students and staff at University College London, were recruited to take part as observers in our user study. All participants had normal or corrected to normal eye sight.

(a) Motion parallax & stereoscopy with vertical viewing angle -10°

(b) Conventional 2D with vertical viewing angle 20°

Figure 6.2: Pictures of the experiment room were taken from different display conditions and vertical viewing angles.

6.2.2.2 Design

The experiment had a 3 perspective conditions $\times$ 3 horizontal viewing angles $\times$ 3 vertical viewing angles $\times$ 35 target positions mixed design, with a within-subjects design for target positions but a between-subject design regarding perspective conditions, horizontal viewing angles and vertical viewing angles.

6.2.2.3 Apparatus and materials

Figure 6.1 and Figure 6.2 show the layout of the experiment room. We arranged small rings as potential target positions. The rings were 1.5cm in diameter, and were placed in a 13 $\times$ 8 grid. Horizontally, top and bottom rows were 13 numbered cards (0 - 12) in a semicircle of radius 100cm at every 15°. Vertically, each column consists of two cards and 6 rings hung from the ceiling with thin thread 10° apart from one another. To improve discriminability, the rings were colour-coded in the following order: black, yellow, green, white, red, and blue. We further arranged 9 large rings to

control participants' eye position for 9 viewing angles by asking them to view the avatar though one of large rings. The viewing distance from participant to avatar position was approximately 170cm.

In the experiment, we created 35 visual stimulii by rotating the avatar's head to look at 7×5 target positions out of 13×8 potential target positions in a prearranged random order (Table 6.1). It is worth noting that the grid of potential target positions was larger than the area of actual target positions, enabling the quantitative investigation of bias in observer perceived target positions. A new target position was given every 10 seconds. Each target position was gazed at only once, amounting to 35 visual stimulii. The most extreme visual stimulii to the outer-most target positions horizontally and vertically were 45° and 20°, respectively. We ensured the avatar's visual stimulus lined up exactly with the centre of corresponding rings.

6.2.2.4 Procedure

Nine groups of participants were used for testing, and each group had three participants. Each group experienced one of three different perspective conditions with one of three vertical viewing angles. Each observer sat at one of the three horizontal viewing angles (see Figure 6.2). Each observer was given a sheet of paper with an empty grid of 35 squares. The avatar reoriented to a new target every 10 seconds. At the same time an audio prompt to the observers instructed them that this was a new target position. Then, observers would judge which target the avatar was gazing at and then write this in the relevant grid square. The experiment took about 6 minutes. Participants received chocolates as compensation.

6.2.3 Result

6.2.3.1 Horizontal error

The primary measurement in our results was the horizontal error in perceiving targets. Any given stimulus i can be defined by a horizontal position (i_h) and a vertical position (i_v). We defined horizontal error of each target (ε_{i_h}) to be the absolute value of a difference between the horizontal position of observer perceived target (t_{oi_h}) and the horizontal position of the actual target (t_{ai_h}), converted to degrees, based on horizontal targets being 15 ° apart from each other:

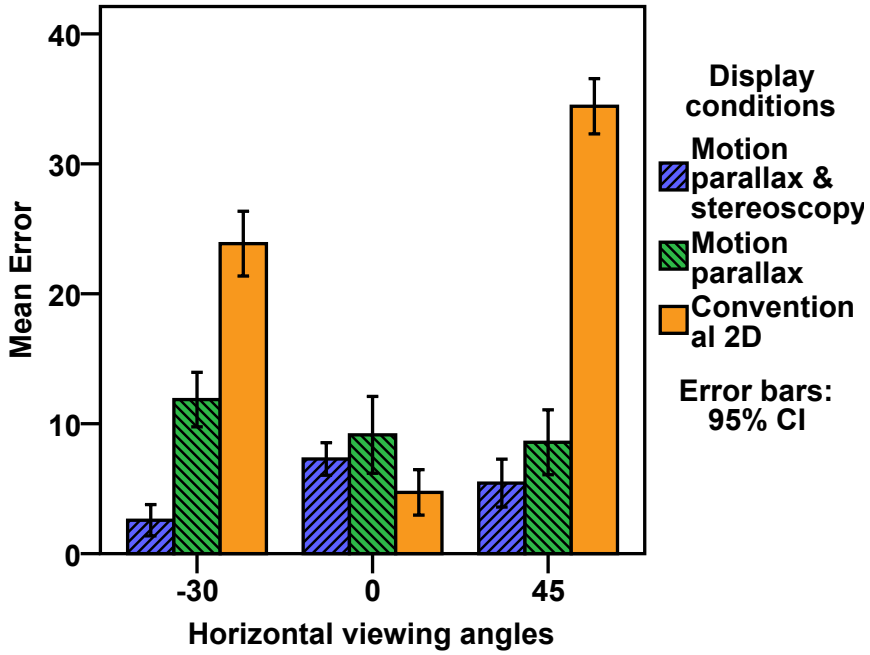


Figure 6.3: The mean horizontal error for each display conditions and horizontal viewing angles.

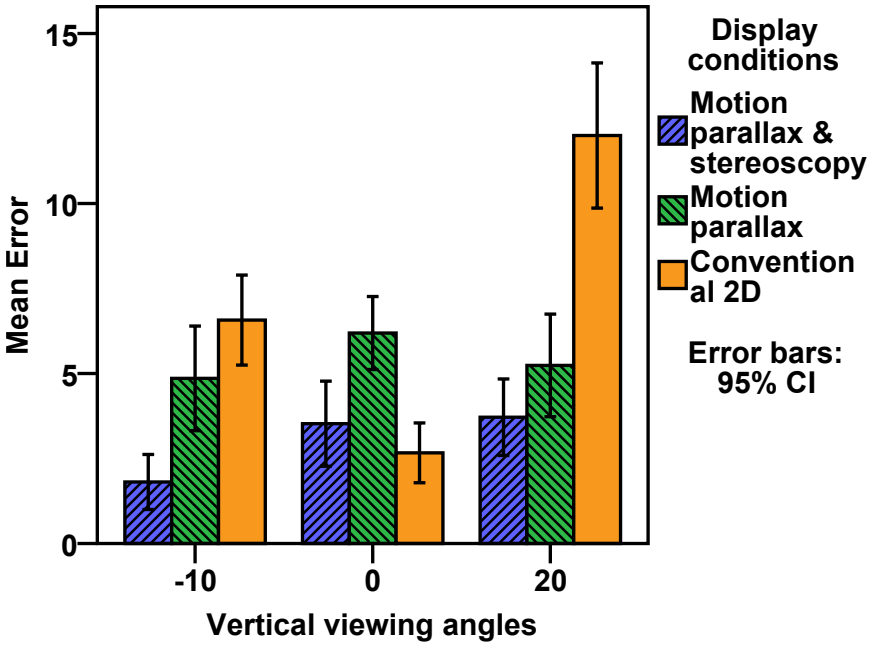


Figure 6.4: The mean vertical error for each display conditions and vertical viewing angles.

$$\varepsilon_{i_h} = |t_{oi_h} - t_{ai_h}| \times 15^\circ \tag{6.1}$$

Figure 6.3 shows the mean horizontal error over all target positions at the three horizontal viewing angles for each of the three display conditions. Overall, the means

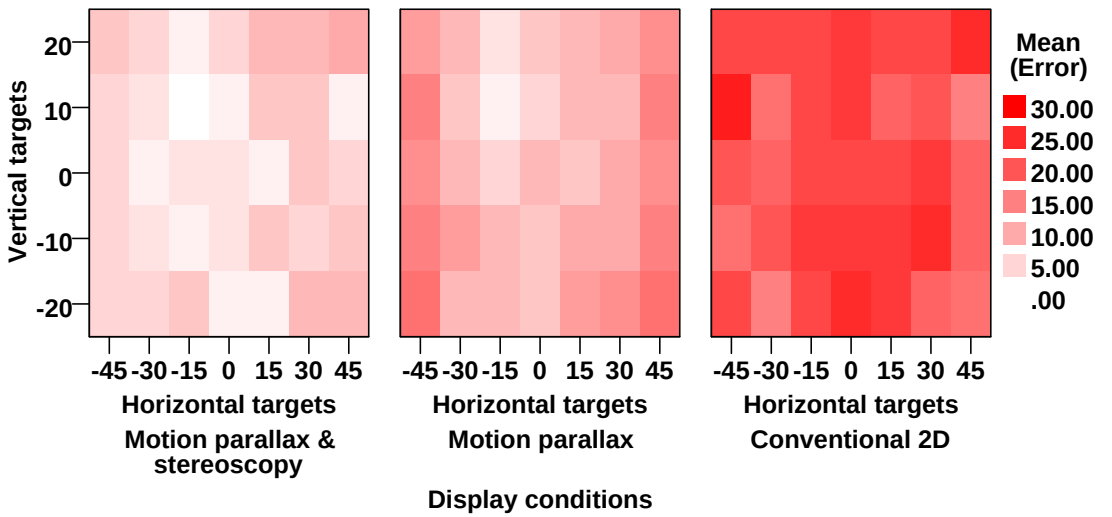


Figure 6.5: Heat maps showing the mean horizontal error for each display condition and target position.

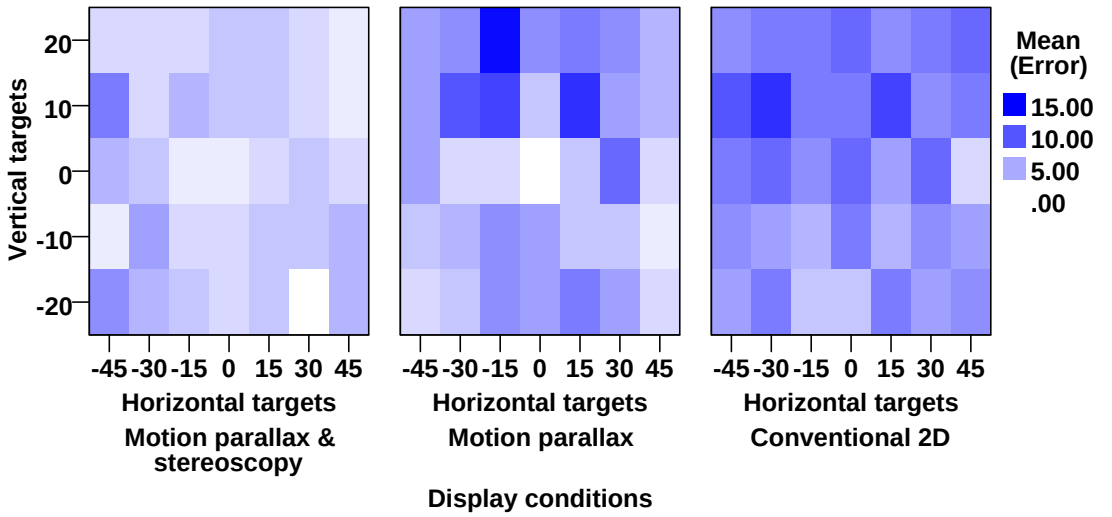


Figure 6.6: Heat maps showing the mean vertical error for each display condition and target position.

of the motion parallax & stereoscopy condition show that it achieved the lowest mean horizontal error. For both the motion parallax & stereoscopy condition and the motion parallax condition, the errors were similar across the three viewing angles, indicating that the viewing angle had little impact. However, for the conventional 2D condition, the errors increased symmetrically as the viewing angle diverged from the central. Figure 6.5 shows the mean horizontal error over all observer's viewing angles for each target positions and display conditions. For the target positions in the motion parallax & stereoscopy condition and the motion parallax condition, the mean horizontal errors are less than 15° (one target error). Interestingly, the errors in the motion par-

allax & stereoscopy condition were more evenly distributed than the motion parallax condition. The motion parallax condition resulted in higher errors when viewing the horizontal edges of the target position grid than the more central locations.

A 3 display conditions $\times$ 3 horizontal viewing angles $\times$ 3 vertical viewing angles $\times$ 7 horizontal target positions mixed design ANOVA was conducted on the horizontal error, with display condition, horizontal viewing angles and vertical viewing angles as between-subjects factors and horizontal target positions as a within-subjects factor. Firstly, the main effect of display conditions was significant, $F(2,108) = 341.029, p < .001$. Bonferroni post-hoc tests revealed significant mean horizontal error differences between each of the display conditions. The motion parallax & stereoscopy condition ($M = 5.095, 95\%\ CI\ [4.219, 5.971]$) gave significantly lower mean horizontal error than the motion parallax condition ($M = 9.857, 95\%\ CI\ [8.981, 10.733]), p < .001$, and the conventional 2D condition ($M = 21, 95\%\ CI\ [20.124, 21.876]), p < .001$. This supports the hypothesis 1a. Secondly, results revealed a significant main effect of horizontal viewing angles, $F(2,108) = 108.166, p < .001$. Bonferroni post-hoc tests revealed significant mean differences between each of the horizontal viewing angles. The mean at viewing angle $0°(M = 7.048, 95\%\ CI\ [6.171, 7.924])$ is significantly lower than viewing angle $-30°(M = 12.762, 95\%\ CI\ [11.886, 13.638]), p < .001$ and viewing angle $45°(M = 16.143, 95\%\ CI\ [15.267, 17.019]), p < .001$. The display conditions $\times$ horizontal viewing angle interaction was significant, $F(4,108) = 146.865, p < .001$, indicating that the error due to viewing angles were different in three display conditions. This supports the hypothesis 2a. Thirdly, we employed Mauchly's test of sphericity to validate our repeated measures factor ANOVAs, thus ensuring that variances for each set of difference scores are equal. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(20) = 70.799, p < .001$), therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .804$). The mean horizontal error differed significantly across horizontal target positions, $F(4.826, 521.216) = 5.148, p < .001$. The display conditions $\times$ horizontal target positions interaction was also significant, $F(9.652, 521.216) = 6.198, p < .001$, indicating that the error due to horizontal target positions was different in three display conditions.

6.2.3.2 Vertical error

We then defined vertical error of each target (ε_{i_v}) to be the absolute value of difference between the vertical position of observer perceived target (t_{oi_v}) and the vertical position of actual target (t_{ai_v}) converted to degrees, based on attention targets being 10° apart from each other:

$$\varepsilon_{i_v} = |t_{oi_v} - t_{ai_v}| \times 10^\circ \tag{6.2}$$

Figure 6.4 shows the mean vertical error over all target positions at the three vertical viewing angles in three display conditions. The interpretations of the results in Figure 6.4 were similar to those in Figure 6.3. Figure 6.6 shows the mean vertical error over all observers' viewing angles for each target positions and display conditions. The heat maps show that the motion parallax & stereoscopy condition has lower mean horizontal error than the motion parallax condition, particularly when viewing the top edge of the target position grid.

A 3 display conditions $\times$ 3 horizontal viewing angles $\times$ 3 vertical viewing angles $\times$ 5 vertical target positions mixed design ANOVA was conducted on the vertical error, with display condition, horizontal viewing angles and vertical viewing angles as between-subjects factors and vertical target positions as a within-subjects factor. Firstly, the main effect of display conditions was significant, $F(2,162) = 45.483, p < .001$. Bonferroni post-hoc tests revealed significant mean vertical error differences between each of the display conditions. The motion parallax & stereoscopy condition ($M = 3.016, 95\%\ CI\ [2.417, 3.614]$) gave significantly lower mean vertical error than the motion parallax condition ($M = 5.429, 95\%\ CI\ [4.83, 6.027]), p < .001$, and the conventional 2D condition ($M = 7.079, 95\%\ CI\ [6.481, 7.678]), p < .001$. This supports the hypothesis 1b. Secondly, results revealed a significant main effect of vertical viewing angles, $F(2,162) = 26.967, p < .001$. Bonferroni post-hoc comparisons indicated the mean vertical error at vertical viewing angle $20^\circ (M = 6.984, 95\%\ CI\ [6.386, 7.583])$ is significantly higher than vertical viewing angle $-10^\circ (M = 4.413, 95\%\ CI\ [3.814, 5.011]), p < .001$ and vertical viewing angle $0^\circ (M = 4.127, 95\%\ CI\ [3.529, 4.725]), p < .001$. However, the mean vertical error at vertical viewing angle 0° did not significantly differ from vertical viewing

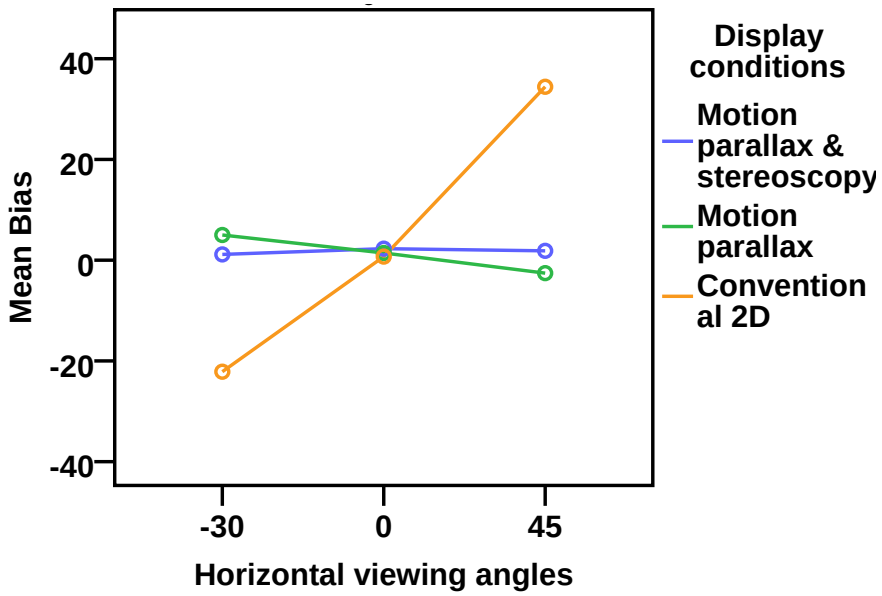


Figure 6.7: The mean horizontal bias for each display conditions and horizontal viewing angles.

angle $-10^{\circ}(p > .05)$. The display conditions $\times$ vertical viewing angle interaction was significant, $F(4,162) = 29.25, p < .001$, indicating that the error due to viewing angles were different in three display conditions. This supports the hypothesis 2b. Thirdly, Mauchly's test indicated that the assumption of sphericity had not been violated $(\chi^2(9) = 8.97, p > .05)$. The mean vertical error differed significantly across vertical target positions, $F(4,648) = 7.189, p < .001$. The display conditions $\times$ vertical target positions interaction was also significant, $F(8,648) = 2.801, p = .005$, indicating that the error due to vertical target positions was different in three display conditions.

6.2.3.3 Horizontal bias

We further investigated whether there was leftward bias or rightward bias in perceiving targets in different display conditions. We defined the horizontal bias of each target (β_{i_h}) to be the difference between the horizontal position of observer's perceived target (t_{oi_h}) and the horizontal position of the actual target (t_{ai_h}) converted to degrees:

$$\beta_{i_h} = (t_{oi_h} - t_{ai_h}) \times 15^{\circ} \tag{6.3}$$

Figure 6.7 shows the horizontal bias at three viewing angles in three display conditions. Positive values indicated leftward biases whereas negative values indicated rightward bias. For both the motion parallax & stereoscopy and the motion parallax

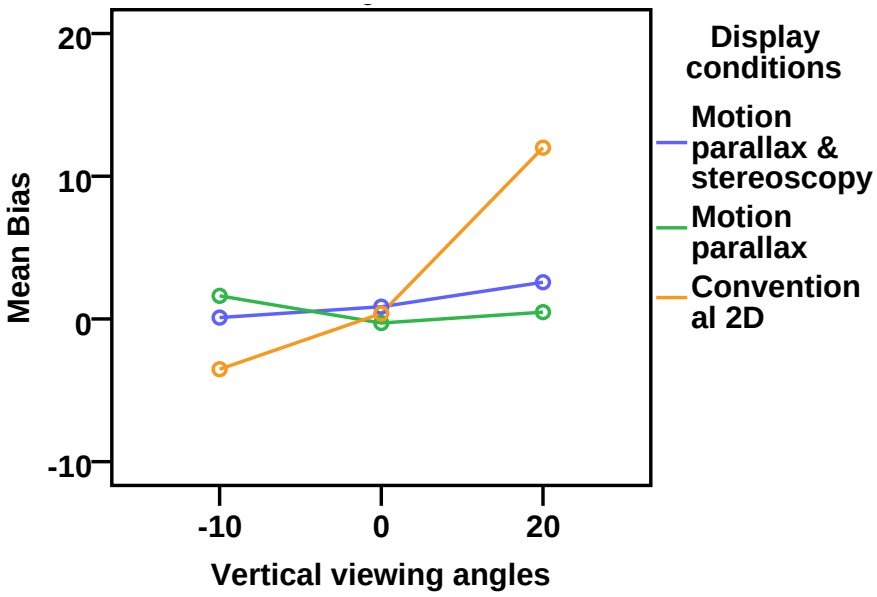
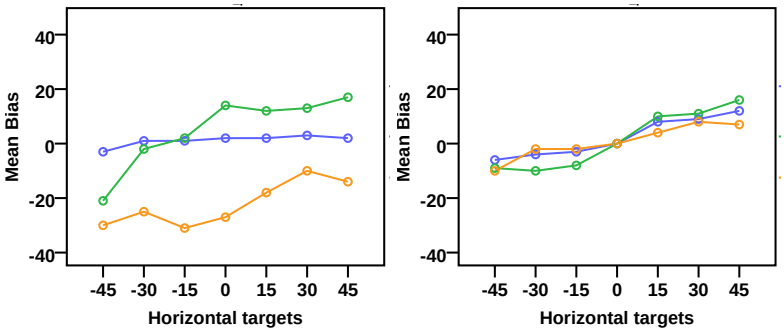
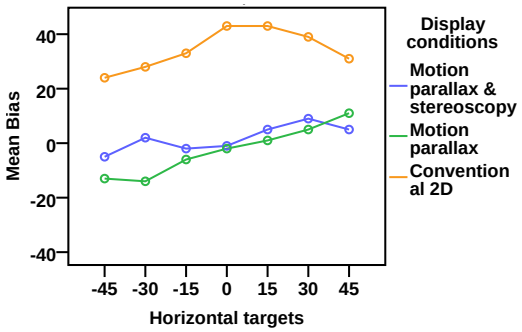


Figure 6.8: The mean vertical bias for each display conditions and vertical viewing angles.



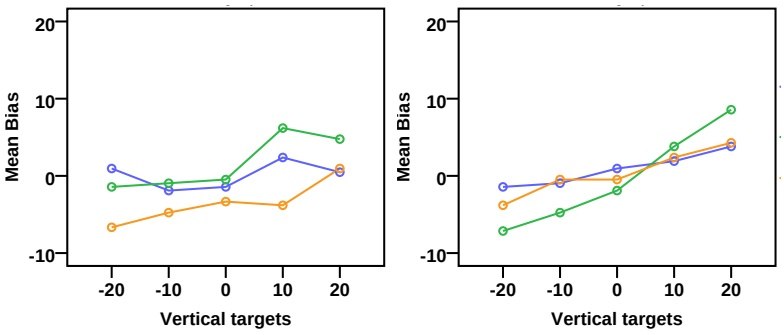
(a) Horizontal viewing angle $-30°$ (b) Horizontal viewing angle $0°$



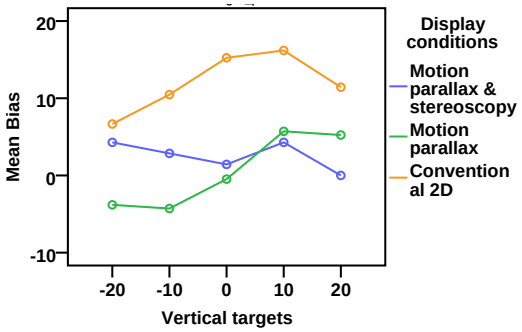
(c) Horizontal viewing angle $45°$

Figure 6.9: The mean horizontal bias for each display conditions, horizontal viewing angles and horizontal target position.

conditions, the mean target bias did not change substantially across different viewpoints. By contrast, for the conventional 2D condition, the biases depended on the observers' horizontal viewing angles. When we consider the target positions in the



(a) Vertical viewing angle $-10°$ (b) Vertical viewing angle $0°$



(c) Vertical viewing angle $20°$

Figure 6.10: The mean vertical bias for each display conditions, vertical viewing angles and horizontal target position.

Figure 6.9, we see the bias doesn't vary with target positions for the motion parallax & stereoscopy condition, however, it increases as the target position gets further away from the observer for motion parallax condition. Considering the horizontal observer at viewing angle $-30°$, we see that for the target position $-30°$, both the motion parallax & stereoscopy condition and the motion parallax condition has similar bias around $0°$; however, for the target position $45°$ the bias increases to $20°$ in motion parallax condition. This overestimation pattern is repeated for all horizontal viewing angles.

A 3 display conditions $\times$ 3 horizontal viewing angles $\times$ 3 vertical viewing angles $\times$ 7 horizontal target positions mixed design ANOVA was conducted on the horizontal bias, with display condition, horizontal viewing angles and vertical viewing angles as between-subjects factors and horizontal target positions as a within-subjects factor. Firstly, the main effect of display conditions was significant, $F(2,108) = 15.068, p < .001$. However, Bonferroni post-hoc tests revealed that the mean horizontal bias in the motion parallax & stereoscopy did not significantly differ from the motion parallax condition, $p > .05$. Secondly, results revealed a significant main ef-

fect of horizontal viewing angles, $F(2,108) = 388.936, p < .001$. Bonferroni post-hoc tests revealed significant mean differences between each of the horizontal viewing angles. Thirdly, Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(20) = 68.76, p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .819$). The mean horizontal bias differed significantly across horizontal target positions, $F(4.914, 530.689) = 125.396, p < .001$. The display conditions $\times$ horizontal target positions interaction was significant, $F(9.828, 530.689) = 11.389, p < .001$. The horizontal viewing angle $\times$ horizontal target positions interaction was significant, $F(9.828, 530.689) = 1.832, p < .001$. The display conditions $\times$ horizontal viewing angle $\times$ horizontal target positions interaction was also significant, $F(19.655, 530.689) = 6.515, p < .001$, indicating that the bias due to horizontal target positions was present differently in three horizontal viewing angles and three display conditions.

6.2.3.4 Vertical bias

Next, we defined the vertical bias of each target(β_{i_v}) to be the difference between the vertical position of observer's perceived target (t_{oi_v}) and the vertical position of actual target(t_{ai_v}) converted to degrees:

$$\beta_{i_v} = (t_{oi_v} - t_{ai_v}) \times 10^{\circ} \tag{6.4}$$

Figure 6.8 shows the vertical bias at three viewing angles in three display conditions. Figure 6.10 shows the vertical bias for each display conditions, vertical viewing angles and horizontal target position. Positive values indicated upward biases whereas negative values indicated downward bias. The interpretation of the vertical bias were similar to those of horizontal bias, but with less effect.

A 3 display conditions $\times$ 3 horizontal viewing angles $\times$ 3 vertical viewing angles $\times$ 5 vertical target positions mixed design ANOVA was conducted on the vertical bias, with display condition, horizontal viewing angles and vertical viewing angles as between-subjects factors and horizontal target positions as a within-subjects factor. Firstly, the main effect of display conditions was significant, $F(2,162) = 13.141, p < .001$. However, Bonferroni post-hoc tests revealed that the mean vertical bias in the motion parallax & stereoscopy did not significantly differ from the motion parallax

condition, $p > .05$. Secondly, results revealed a significant main effect of vertical viewing angles, $F(2,162) = 79.521, p < .001$. Bonferroni post-hoc comparisons indicated the mean vertical bias at vertical viewing angle 20° significantly differs from vertical viewing angle $-10°$, $p < .001$ and vertical viewing angle 0°, $p < .001$. However, the mean vertical error at vertical viewing angle 0° did not significantly differ from vertical viewing angle $-10°$, $p > .05$. Thirdly, Mauchly's test indicated that the assumption of sphericity had not been violated ($\chi^2(9) = 13.571, p > .05$). The mean vertical bias differed significantly across vertical target positions, $F(4,648) = 37.908, p < .001$. The display conditions $\times$ vertical target positions interaction was significant, $F(8,648) = 9.108, p < .001$. The vertical viewing angle $\times$ vertical target positions interaction was significant, $F(8,648) = 3.826, p > .05$. However, the display conditions $\times$ vertical viewing angle $\times$ vertical target positions interaction was not significant, $F(16,648) = 1.562, p > .05$.

6.3 Discussion

Results from this experiment confirmed our hypotheses. We found that participants performed with the lowest error when interpreting the avatar's gaze direction in the motion parallax & stereoscopy condition, followed by the motion parallax condition, and then the traditional 2D condition. This is consistent with Kim et al.'s previous findings in 3D video communication [50].

The poor performance of the traditional 2D condition was expected because the head is always rendered from a front perspective. The only position with the correct perspective would be the observer at centre where the front perspective correlates to that observer's perspective. From the rest of viewing positions, the observers would be experiencing from the Mona Lisa gaze effect. They would perceive the gaze direction as if they were standing straight in front of the display. Thus, they would see the gaze in a relative rather than an absolute manner. As expected, Figure 6.9 and Figure 6.10 show that the curves of the traditional 2D condition maintain a similar shape, but are shifted depending on observer's perspective. This parallels the previous findings [3, 71] in 2D video condition.

For the comparison the motion parallax condition and the motion parallax & stereoscopy condition, we found the differences in vertical and horizontal errors were sta-

tistically significant. However, the differences in vertical and horizontal bias were not statistically significant. This suggested that motion parallax alone could reduce the shifting bias discussed above.

Figure 6.9 and Figure 6.10 show that the overestimation pattern in the motion parallax condition is interesting. They indicated the addition of stereoscopy could reduce an overestimation of the deviation of avatar's gaze, thus further improving the observers' ability to identify more correct targets. This was also backed up by results from the vertical and horizontal errors. An analysis of the heat maps in Figure 6.5 and Figure 6.6 show that observers performed with higher level of error when viewing the edges of the target grid than the more central locations in the motion parallax condition. This effect appears very reliable and this means that it may be possible to model and thus predict the distortion.

We also investigated judgments of vertical direction of gaze. Figure 6.9 and Figure 6.10 show that the magnitude of the shifting bias in 2D condition and the overestimation pattern in the motion parallax condition are smaller in vertical direction comparing to horizontal direction. This discrepancy in results between judgments of horizontal and of vertical gaze reflects the asymmetric sensitivity of users when perceiving avatar's head outline. This is supported by the previous findings [133] that the perceived direction of gaze can be influenced by deviation of the head profile from bilateral symmetry, and deviation of nose orientation from vertical.

6.4 Chapter summary

In this chapter, we ran an experiment to demonstrate that the random hole display can convey gaze relatively accurately, particularly for group conferencing. We further investigated the effects of reproducing motion parallax and stereoscopic cues in telepresence in both horizontal and vertical directions. We provided detailed reasons for the improvement of our system in conveying gaze. We compared three different conditions: conventional 2D, motion parallax, and motion parallax & stereoscopy across nine varying viewing angles. Results show that the presence of both motion parallax and stereoscopic cues significantly improved the accuracy with which participants were able to assess the avatar's gaze in both horizontal and vertical directions. This demonstration motivates the further study of novel display configurations and suggests

parameters for the design of teleconferencing systems.

Chapter 7

Experiment: Trust in spherical avatar telepresence system

"We're never so vulnerable than when we trust someone but paradoxically, if we cannot trust, neither can we find love or joy" Walter Anderson. When people need to establish trust at a distance, it is advantageous for them to use rich media to communicate.

Trust is an important factor in many facets of our lives. In business settings, trust is required in order for people to work together effectively. Without trust, they will not share information openly, and transactions must be carefully contracted and monitored to prevent exploitation. They may also change the nature of collaborations to avoid the need for close coordination or may simply avoid collaborating with others altogether, thus limiting their productive capacity. But if higher degrees of trust can be established, people can work more efficiently, and adapt more quickly to changing situations.

As reviewed in section 2.4.2, there is a growing body of literature on how computer-mediated communication systems affect trust formation. In this chapter, we investigated the influence of display type and viewing angle on how people place their trust during avatar mediated interaction. In our experiments, participants were required to attempt to answer thirty difficult general-knowledge questions. For each question, participants could ask for advice from one of two advisers. Unknown to the participants, one was an *expert* who responded with mainly correct information, and the other was a *non-expert* who provided mainly incorrect information. We measured participants' advice seeking behavior as an indicator of their trust in the adviser. The first experiment explores how interpersonal cues of expertise presented on two identical flat displays with different viewing angle affect trust. The results demonstrate that partic-

ipants were able to discriminate correct advice, but their sensitivity to correct advice decreased at off-center viewing angles. The second experiment compares two display types by investigating how people place their trust. Balanced over participants, the expert appeared either on the sphere or on the flat display. We found most participants preferred seeking advice from the expert, but we also found a tendency for seeking advice from the adviser on the spherical display instead of flat display, in particular when viewed from off-center directions.

7.1 Evaluation design: advice seeking behavior

Through two experiments, we investigated how display type affects trust. Our first experiment (E1) explored the effect of viewing angle on trust in traditional flat displays, and provided a benchmark by which to measure the spherical display. Our second experiment (E2) investigated the impact of the spherical display given that it could faithfully reproduce the actor's gaze at all viewing directions.

We modeled our experiments on a user-adviser relationship [92], a widely used research paradigm in social psychology. Participants were asked to answer thirty difficult general-knowledge questions and they received chocolates depending on their performance. We gave participants two advisers presented on two teleconferencing displays. Unknown to participants, the two advisers are with different levels of expertise. Additionally, the spatial arrangement of participant-to-displays was varied over the course of the experiment, thereby manipulating participants' viewing angle of the advisers. Advice was free, but only one adviser could be asked per question.

We measured participants' advice seeking behavior under risk as an indicator of trust in the adviser. People generally decide to trust others when facing situations involving risk and uncertainty [31, 65]. Uncertainty arises from the fact that the participants cannot directly observe the two advisers' ability (e.g. expertise) and motivation (e.g. desire to deceive). They need to infer those from interpersonal cues, as the questions were extremely difficult. When recording the non-expert clips, the actor exhibited less direct eye contact and less confident facial expression. When recording the expert clips, the actor exhibited confidence through more positive facial expression, such as smiles and eye contact. In our experiments, viewing angles and display types influence those interpersonal cues. Seeking advice from one adviser in preference over the other

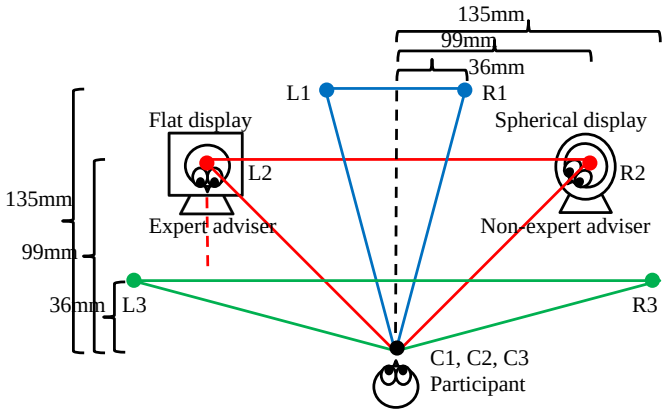


Figure 7.1: Schematic layout of experiment setup. L1, R1 & C1; L2, R2 & C2 and L3, R3 & C3 are three participant-to-displays spatial arrangements. C1, C2 and C3 are participants' seating positions which are 75°, 45° and 15° relative to display, respectively. Also see Figure 7.2 and Figure 7.3.

could be an indication of trust in that adviser, because receiving poor advice carried the risk of missing out better advice and therefore the participant was less likely to get the correct answer.

(a) Photo taken at left side of participant seated at C1

(b) Photo taken behind the participant seated at C2

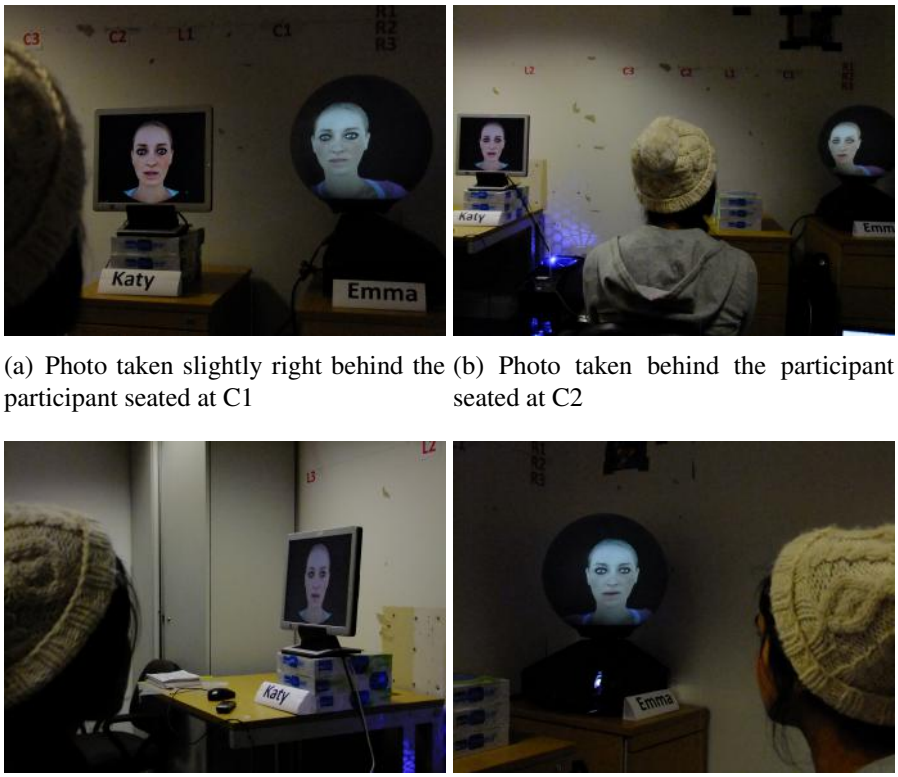
(c) Photo taken slightly right behind the participant seated at C3

Figure 7.2: Picture of E1 room taken from different perspective relative to the participant seated at different seat positions. see Figure 7.1 for seat positions.

7.1.1 Apparatus and materials

7.1.1.1 Questions

For E1 & E2, we used 30 questions and answers and a transcript of advice from previous research on trust in a human adviser [92]. Those questions are difficult general knowledge questions, to minimize effects of participants' prior knowledge. Examples of questions that were included are 'Which New York Building featured a mural depicting Lenin?' and 'Which one of these is a coastal city in North Korea?'. Based on the pre-test results, the mean probability for pre-testers giving a correct answer was .31



(a) Photo taken slightly right behind the participant seated at C1

(b) Photo taken behind the participant seated at C2

(c) Photo taken at right side of participant seated at C2

(d) Photo taken at left side of participant seated at C2

Figure 7.3: Picture of E2 room taken from different perspective relative to the participant seated at different seat positions. see Figure 7.1 for seat positions.

(SD = .11). This value was only marginally above chance (.25), indicating that very difficult questions had been picked.

7.1.1.2 Expertise

The non-expert and expert advisers were created by recording advice from the same individual before and after training, respectively. The same animations are used in both experiments. We used Faceshift to simultaneously record the actor's performance including voice and blendshape weight vectors that drive the avatar's facial expression. Then, we synchronously replayed both audio and facial expression on the display. The expert and non-expert advisers only differed in the ratio of correct to incorrect advice and in their cues to confidence about the answers. As each time the observer only had access to one of the advisers, they were unaware that both advisers were in fact the same individual recorded at different levels of expertise. For the non-expert adviser, the proportion of correct (i.e. confident) advice was 0.36. For the expert adviser, the

No.	Statement
1	Adviser was very friendly
2	I am pleased with adviser
3	I trusted adviser's advice
4	I enjoying playing with adviser
5	I would like to meet adviser face to face
6	Adviser gave good advice
7	Adviser was certain about the answer
8	I liked adviser
9	I relied mostly on adviser's advice

Table 7.1: Statements for post-experimental assessments of the adviser.

proportion of correct (i.e. confident) advice was 0.80. Two incorrect (and less confident) pieces of advice from the untrained recording were added to the expert, in order to avoid artificial perfection.

Note that the system as designed and built is a realtime collaborative system that can connect a remote room to a local room. For the purposes of our controlled experiment we used pre-recorded clips.

7.1.1.3 Display Type

The participant observes the pre-recorded avatar video clips on two displays. We used two flat displays in E1, whereas one flat display and one sphere display in E2. For the flat display, a conventional PC screen was used with a resolution of 1024×768 pixels. For the sphere display, with perspective-correct ray traced imagery, the participant perceives the avatar to be situated inside the display and looking at him or her. We ensured the avatars' apparent sizes on sphere and flat display were the same (20 cm in height).

7.1.1.4 Seat Position & viewing angle

For both experiments, we arranged the two displays and participants' seat positions at vertices of three isosceles triangles with base angle of 75°, 45° and 15° for three different seat positions (see Figure 7.1). The legs for all those three isosceles triangles, which is the distance between participant and display, were maintained the same at 140 cm. We ensured that the vertical alignment of the eye level of viewers and the eye level of the avatar of the actor on the two displays were the same.

7.1.1.5 Incentives & risk

For both experiments, the number of chocolates that participants received was linked to the number of correctly answered questions. The number of chocolates varied between one and six.

7.1.2 Measurement Instruments

7.1.2.1 Task Performance Measure

The measure of advice seeking was defined as the proportion of one adviser being asked out of the total number of times advice was sought by a participant. As each participant had two advisers, but could only choose one of them for advice on each question, the following relationships hold: expert advice seeking $= 1 -$ non-expert advice seeking, and one display advice seeking $= 1 -$ the other display advice seeking.

7.1.2.2 Post-Questionnaire

Participants were presented with the post-experimental questionnaire with 9 statements (see Table 7.1) eliciting their subjective assessment of the two advisers, with 4 items measuring trustworthiness (Statement 3, 6, 7 & 9) and 5 items measuring enjoyment (Statement 1, 2, 4, 5 & 8). Agreement with the statements was elicited on 7-point Likert scales with the anchor 1 (Strongly disagree) - 7 (Strongly agree).

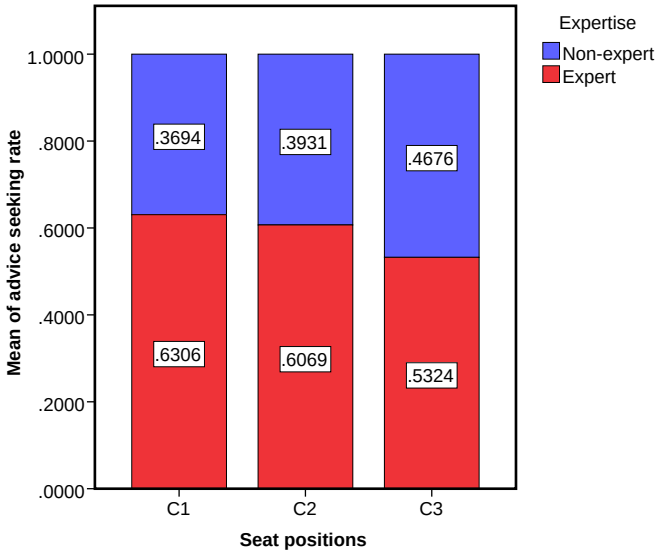
7.1.2.3 Open question

We asked each participant to write down his or her comments with a final open question: "For each round of games, please describe how you decided which adviser to rely on". The purpose of this open question was to help explain some observed events during the game and to guide future research.

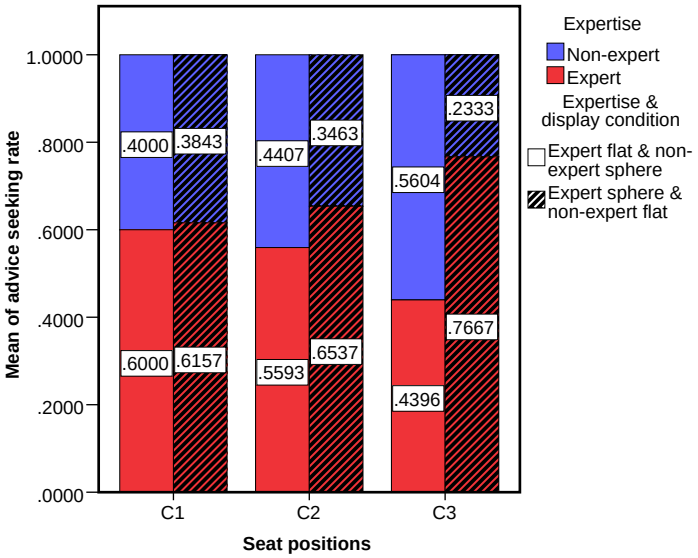
7.2 Experiment 1

7.2.1 Hypotheses

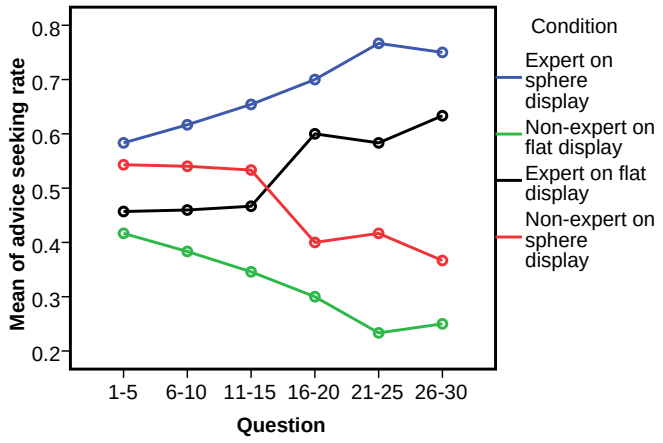
We expect participants to seek more advice from the expert adviser than the non-expert adviser. We further expect that the more the seat position diverges from the central viewing position, the worse the observer will be able to discriminate between trustworthy and less trustworthy advisers. This is because the observer cannot look straight into the display and the slight visual spatial degradation will reduce observer's ability to



(a) E1: Advice seeking across seat positions.



(b) E2: Advice seeking across seat positions.



(c) E2: Advice seeking over time.

Figure 7.4: Results of E1 & E2: task performance measure. see Figure 7.1 for seat positions.

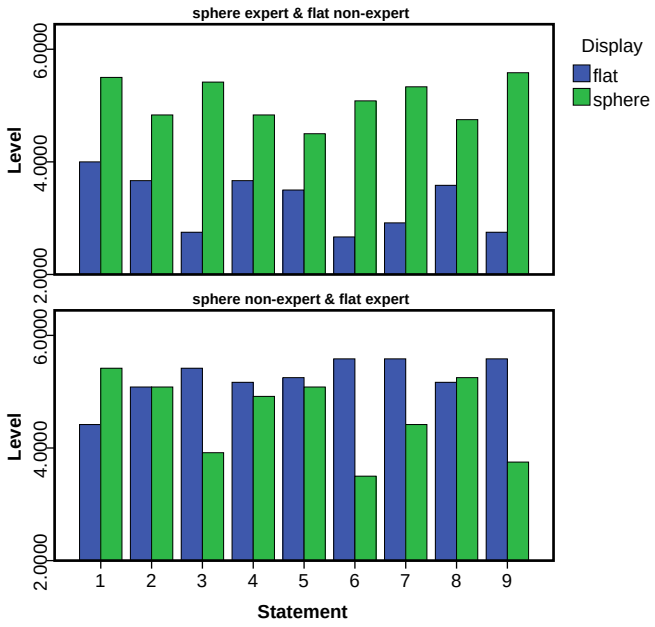


Figure 7.5: Post-experimental assessments of the advisers.

discriminate [44].

7.2.2 Method

Twelve participants (6 male), students and staff at University College London, were recruited to take part as observers in E1. The median age was 21.75 (SD = 3.20).

E1 had a one-way within-subjects design (see Figure 7.2(a) to Figure 7.2(c)). Every participant took part in the experiment at 3 different seat positions (C1, C2 and C3). The order of the answer options (A-D) of questions was randomized. The expertise and the participant's 3 different seat positions order were counterbalanced, in order to reduce any confounding influence of the experiment environment such as lighting conditions and the orderings such as learning effects or fatigue.

Prior to starting the assessed part of the experiment, each participant completed a training round that consisted of easy questions where both advisers gave identical and correct advice. Then, participants answered 10 assessed questions in each round. The participant could ask for advice before answering each question. For each question, participants could ask for advice from one of two advisers without knowing of the adviser's expertise. After each participant played one round at one seat position, the participant moved to another seat position. This process repeated for three different seat positions. Each participant had the same two advisers (Emma and Katy) for the whole

study. After completing all rounds they were presented with the post-experimental questionnaire and an open question. Finally, the participants were compensated with chocolates based on their performance. The experiment took about 30 minutes.

7.2.3 Results

In E1, participants sought advice on 29.33 out of 30 questions (97.78%) over 3 rounds. 7 participants (58.33%) sought advice in every question. There was no cost associated with seeking advice. Figure 7.4(a) shows that the experts (red bar) were chosen more often than non-experts (blue bar) for all three seat positions. However, from seat position C1 to C3, the expert advice seeking rate dropped off whereas non-expert advice seeking rate increased. We interpret this to indicate the decrement of sensitivity for cues of expertise.

A one-way repeated measures ANOVA was conducted to compare the effect of the expert advice seeking rate in 3 seat positions (C1, C2 or C3) conditions. There was a significant effect of seat positions, $F(2,22) = 6.356, p < .01$. Three paired samples t-tests were used to make post hoc comparisons between conditions. When we did three paired samples t-tests, we increased our chances of finding a significant result when one did not exist. Instead of using the value .05 to decide if we had reached statistical significance, we would instead use the value .017 $(= .05/3)$ as the cut off. A first paired samples t-test indicated that there was no significant difference between C1 $(M = 63.06\%, SD = .063)$ and C2 $(M = 60.69\%, SD = .091)$ conditions; $t(11) = .945, p = .365$. A second paired samples t-test indicated that there was a significant difference for C1 and C3 $(M = 53.24\%, SD = .088)$ conditions; $t(11) = 3.457, p = .005$. A third paired samples t-test indicated that there was no significant difference between C2 and C3 conditions; $t(11) = 2.304, p = .042$. The expert advice seeking rate at C3 is significantly less than C1. This suggests that further the seat position aside from the central position, the more difficulty the observer had in identifying the expert. This supports our first hypothesis.

7.3 Experiment 2

7.3.1 Hypotheses

7.3.1.1 Hypothesis 1

We expected participants to seek more advice from the expert adviser than the non-expert adviser. By introducing the spherical display, we expected that the observer's sensitivity to cues of trustworthiness to remain stable for all seat positions, as it conveys the same amount of information for all directions.

7.3.1.2 Hypothesis 2

We expected that the flat display will result in less trust compared to the sphere display. In other words, bias will occur when advice is preferred due to its display mode, irrespective of expertise. We further expected a negative bias towards the flat display representation will be found at off-center viewing angles, due to the loss of eye contact.

7.3.2 Method

Twenty-four participants (12 male) took part in E2. The median age was 21 (SD = 2.30). Participants had not previously interacted with advisers.

E2 is similar to E1, except that instead of presenting two advisers on two identical flat displays, we presented one on sphere display, and the other on flat display. E2 had a 2 display modes (*Expert is sphere display* vs. *expert is flat display*) $\times$ 3 seat positions mixed design, resulting in 2 between-subject conditions with 12 participants each (see Figure 7.3(a) to Figure 7.3(d)). In each between-subject condition, two advisers were available. Depending on the display mode, either the sphere display or the flat display adviser gave expert advice, while the other gave non-expert advice. The two display positions (left-right) were counterbalanced by switching around the sphere and flat displays. To moderate the effect introduced by evaluating a novel type of display, we asked each participant to complete a practice round prior to starting the assessed part of the experiment.

7.3.3 Results

7.3.3.1 Display type & viewing angle

In E2, participants sought advice on 29.21 out of 30 questions (97.36%) on average. 15 participants (62.5%) sought advice in every question. Figure 7.4(b) shows that the experts (red bar) were chosen more often than non-experts (blue bar) for all three seat positions. The overall expert advice seeking rate (expert on sphere display + expert on flat display) were 60.79%, 60.65% and 60.31% at the seat position C1, C2 and C3, respectively. This indicated that the overall expert advice seeking rate remained the same among three seat positions, which were approximately 20% higher than overall non-expert advice seeking rate. Figure 7.4(b) also shows that a preference for choosing sphere display increased from seat position C1 to C3, while decreased in the flat display condition. The overall sphere display seeking rate (expert on sphere + non-expert on sphere) were 50.78%, 54.72% and 66.35% at the seat position C1, C2 and C3, respectively. Sphere display advice seeking rate was higher in seat position C3. We note that for seat position C1, the flat display and sphere display were chosen equally often. This is expected as the seat position only slightly diverges from the front and the faces of two advisers can be seeing similarly on both display types.

A 2 (display: flat vs. sphere) $\times$ 3 (seat positions: C1, C2 or C3) repeated measures ANOVA was conducted on the expert advice seeking rate, with display as a between-subjects factor and seat positions as a within-subjects factor. This revealed a significant main effect of display, $F(1,22) = 13.757, p < .01$, indicating that expert advice seeking rate was significantly higher for sphere display. There was no significant main effect of seat positions, $F(2,44) = .011, p > .05$, indicating overall expert advice seeking rate at different seat positions were not statistically significant different from one another, thus further supporting our first hypothesis. However, the display $\times$ seat position interaction was significant, $F(2,44) = 11.745, p < .001$, indicating that expert advice seeking rate due to seat position was presented differently in sphere and flat display conditions. This supports the second hypothesis.

We further investigated sphere display non-expert advice seeking rate at three different seat positions (unshaded blue bar in Figure 7.4(b)). The non-expert advice seeking rate $< .5$ would provide evidence for users' ability to discriminate between expert

and non-expert advisers, whereas the value $> .5$ would be a sign of bias outweighing discrimination. Based on a one-sample t-test, the sphere display non-expert advice seeking rate at seat position C1 is significantly below .5, $t(11) = -2.582, p < .05$. There is also some indication at seat position C2, $t(11) = -2.111, p = .058$. However, no such effect is presented at seat position C3, $t(11) = 1.781, p > .05$, indicating that a bias towards sphere display is interfering with users' ability to discriminate, thus supporting the second hypothesis.

We then analyzed how participants' advice seeking behavior changes over time. Figure 7.4(c) presents the mean advice seeking rate of every five questions in chronological order. The choice to seek advice from a specific adviser could be expected to depend upon the information accumulated from previous pieces of advice. It was thus assumed to be relatively arbitrary in initial interactions. Participants increasingly sought advice from the expert as they gained experience with the advisers, but there is a bias towards the sphere display. This gives us further evidence for the second hypothesis.

7.3.3.2 Post-Questionnaire

Figure 7.5 shows the result of the participants' self-reports. In the expert on sphere display condition, the statements measuring ability (Statement 3, 6, 7,& 9) were higher for the sphere display; and in the expert on flat display condition, those statements were higher for the flat display. This indicated that participants were able to identify the trustworthy adviser. However, statements measuring enjoyment (Statement 1, 2, 4, 5,& 8) showed similar or higher level of score for sphere display, even in the expert on flat display condition. This indicated that using the sphere display could increase social presence.

The responses to each statement item given by all the participants were averaged to create an aggregate response. We calculated Cronbach's alpha as the reliability test. The questionnaire measured four subscales: trustworthiness of the sphere display adviser (4 items, $\alpha = .893$), trustworthiness of the flat display adviser (4 items, $\alpha = .96$), enjoyment of the sphere display adviser (5 items, $\alpha = .807$), and enjoyment of the flat display adviser (5 items, $\alpha = .932$). We then analyzed the post-experimental assessments of the advisers by comparing each participant's rating of the sphere display

adviser to that of the flat display adviser, irrespective of the expertise of each adviser. Significant differences in the post-experimental assessment (see Table 7.1) between sphere display and flat display adviser are thus indicators of bias on one subscale for one specific display type. Two paired-samples t-test were conducted to compare the ratings of the trustworthiness and enjoyment in sphere display and flat display conditions. Notable bias was found for enjoyment, sphere display rated as being friendlier than flat display, irrespective of expertise, $t(23) = -2.228, p < .05$.

7.4 Discussion

We compared the advice-seeking rate at three seat positions between E1 and E2. E1 utilized two flat displays, with results demonstrating that participants' sensitivity to correct advice decreased at the far off-center viewing positions (C3). By introducing the spherical display in E2 that was able to preserve correct gaze direction and a simple pseudo-3D experience by providing perspective-correct rendering at all viewing angles using non-planar surface, we found participants' ability to discriminate remained stable at all viewing positions.

From participants' behavioral measures, we found that participants mostly chose expert advice in both flat and sphere display representations. This indicates that participants were able to discriminate between experts and non-experts, and accordingly, distributed more trust to the expert. However, there was also evidence that display representation can interfere with participants' ability to discriminate effectively. The sphere display produced a higher rate of advice seeking compared with the flat display. This behavioral manipulation emerged at off-center viewing positions and increased as the viewing position became more extreme. At the most extreme viewing position (C3), the rate of advice-seeking from the avatar displayed on the sphere was significantly greater than that sought from the avatar shown on the flat display. The preference for seeking advice from the avatar on the sphere display almost matched the preference for choosing expert advice, despite participants generally knowing on which display the expert was positioned. This negative bias towards the flat screen at off-center viewing angles in avatar-mediated communication parallels a similar finding by Nguyen et al. [72] in video-mediated communication. In that study, they examined the effects of spatial faithfulness on trust formation in a cooperative investment task.

They found the spatial distortions of traditional flat display negatively affect trust formation patterns. The finding that trust can be easily and significantly manipulated in mediated interaction by adjusting display viewing angle has significant implications for telecommunication in general. We plan further investigation on this topic, with our next step being to quantitatively evaluate gaze of participants using eye tracking and introduce another between-subject condition (sphere display expert vs. sphere display non-expert) to further explore this finding.

Our post-experimental open question further supports our findings. It should be noted that Katy was the adviser on flat display and Emma was the adviser on sphere display. In the expert is sphere condition, one participant stated "It is difficult to see Katy speak and look at her expressions while she answered, I could not feel good to communicate with Katy. Thus, I chose Emma more times." Regarding viewing angle, another participant stated "I was sitting facing them directly rather than an angle with Emma, the more I felt they were reliable." In the expert is flat condition, one participant expressed "Emma's eye gives a supporting feeling, but Katy's voice is more confident. Katy seems always certain about the answer, but Emma seems to tell what she knows." Participants' answers also show that there were other factors influencing their decision making, with one stating "I got a fully confident answer by myself and Katy also told me the matched answer, so I tended to ask her more.", and another stating "The longer time I spend with Katy and Emma, I figure out who knows more answers. But sometimes I still need to double check."

7.5 Chapter summary

The two experiments reported in this chapter aimed to investigate the influence of display type and viewing angle on how people place their trust during avatar-mediated interaction. By monitoring advice seeking behavior, our first experiment demonstrates that if participants observe an avatar at an oblique viewing angle on a flat display, they are less able to discriminate between expert and non-expert advice than if they observe the avatar face-on. We then introduce a novel spherical display and a ray-traced rendering technique that can display an avatar that can be seen correctly from any viewing direction. We expect that a spherical display has advantages over a flat display because it better supports non-verbal cues, particularly gaze direction, since it presents a

clear and undistorted viewing aspect at all angles. Our second experiment compares the spherical display to a flat display. Whilst participants can discriminate expert advice regardless of display, a negative bias towards the flat screen emerges at oblique viewing angles. This result emphasizes the ability of the spherical display to be viewed qualitatively similarly from all angles. Together the experiments demonstrate how trust can be altered depending on how one views the avatar.

Chapter 8

Conclusions

With collaborative efforts increasingly spanning large distances and the economic and environmental impact of travel becoming increasingly burdensome, telepresence techniques are becoming increasingly widespread. The goal of any computer-mediated communication system is to allow geographically separated parties to meet effectively.

With the understanding that nonverbal cues can play a significant role in communication, this thesis analyzed the mechanisms required for effective use of nonverbal cues, particularly gaze, with respect to how current teleconferencing systems fail to support these mechanisms. We introduced four novel telepresence displays (Chapter 3). The follow up with studies demonstrated the affordances of our systems (Chapter 4 to Chapter 7).

This closing chapter summarises the work presented in this thesis. Firstly, the affordance, limitations, & applications of each telepresence system, and the findings of each related experiment are recounted. This is followed by the holistic conclusion, relating back to the research problems and contributions established in Chapter 1. Finally, potential direction for future work are established.

8.1 Spherical video telepresence system

For the first system, we developed a novel spherical video telepresence system in order to give an observer some of the advantages of meeting face-to-face without the disadvantages of traveling. This display offers a 360° view whereas a flat display is only visible from the front. By using a surrounding camera array, we allow a single principal observer to accurately tell where the actor is looking from multiple observing positions at all angles. The captured video is projected from the bottom of spherical display,

which successfully maintains the gaze fidelity without reducing the quality of the video and complexity of the display system (e.g., using a two-way mirror). This motivates further development of video conferencing systems that exploit multiple cameras and non-planar displays.

The spherical video telepresence system could be used in a teaching scenario or a tele-surgery application where a remote person instructs a local user. The local user could perceive precise spatial information from any viewpoint in the room whereas flat displays are only visible from the front. Our current system is used for asymmetric conversations, however, systems using similar principles could be configured to support symmetric conversations, by arranging camera arrays that are denser but further from the users.

An interesting question is the potential support for multiple viewers. The evaluation of the secondary positions in the first experiment, the three "three point hat" graphs demonstrated that the gaze cues are only preserved for the principal observer. This is because the position of the observer is needed in order to render the head correctly for that perspective. The spherical display could be made for multiple viewers. The inflated display mode of SphereAvatar [78] supported multiple viewers in avatar-mediated teleconferencing. For video mediated teleconferencing, we could project a whole head by using the similar idea proposed by [46] in the one-to-many 3D video teleconferencing system.

In the spherical video telepresence system, the video texture is projected on a sphere. An alternative approach would have been to project onto an ellipsoid or a more "head-shaped" object than a sphere, however, this would have worked for head rotations around the vertical axis while the projection would be severely distorted for rotations around other axes.

In addition, it would be interesting to investigate novel rendering methods to avoid the steep drop in accuracy when the observer is not aligned with the cameras by interpolating between videos. Furthermore, it would also be interesting to investigate less constrained positioning of the cameras and different eye-lines. As noted, although the experiment used recorded data, the system can run in a live, automatic camera switching mode and thus it would be interesting to investigate how users utilize movement to control the video.

We are the first to compare situated display and flat display in preserving gaze direction. We have demonstrated that the sphere display preserves the accuracy of observing actor's gaze direction, even at extreme seat positions. This may be due to the ability of sphere displays to produce a correct view. Furthermore, we proposed two linear models for predicting the spatial distortion introduced by misalignment of capturing cameras and observer's viewing angles. Therefore, we might be able to correct for this distortion in future display configurations.

8.2 Cylindrical video multiview telepresence system

For the second system, we have presented a novel cylindrical video telepresence system for video conferencing. The highlights of this system are as follows. Firstly, the cylindrical display offers a wide field of view whereas flat displays are only visible from the front. Secondly by using a surrounding camera array, a projector array and a multiview screen, we are able to transmit the remote person to multiple observers gathered around the cylindrical display, maintaining accurate cues of gaze direction.

A similar cylindrical multiview display could also use a very dense projector array covering 360°, thus supporting a large number of viewpoints from any directions without introducing crosstalk and reducing resolution. As cameras and projectors are now becoming very cheap, the low cost and ease of setup make this an interesting platform for next generation video conferencing.

8.3 Random hole autostereoscopic multiview telepresence system

For the third system, we have presented a ray-traced view-dependent rendering method to represent the remote person as a virtual avatar on the random hole display. It offers a number of capabilities that are not found in most existing autostereoscopic displays, including display for multiple users in arbitrary viewing positions. The observers maximum viewing angle depends on the LCD panels viewing angle. The low cost and ease of setup make this system an interesting platform to simulating scenarios that require multiple simultaneous stereo views from arbitrary positions.

We used the SIPS type display and the maximum viewing angle is at least 70 degrees in each direction. Although the random hole type displays have a limited spatial

resolution (see section 3.4), on our display the different views are easily distinguished. Figure 3.21 is a set of stereo pair images, showing the autostereoscopic image quality provided to three users. Additionally, all participants in our experiment confirmed that they are able to clearly tell where the avatars eyeball is actually looking. In the future, we expect to use brighter and higher-density LCD/LED panels or high-resolution multiple projector systems as display surfaces to further improve image quality.

With our current demonstration we are using a ray-traced avatar head. Although the animation we have used in the experiment is simple, the software system supports a fully animated head with eye movement and facial expression, using Faceshift® with Microsoft Kinect™ to obtain the remote person's eye movement and facial expression in realtime.

There are several routes for development to support different conversation scenarios. Firstly, our current system can be used for asymmetric conversations. This setup could be mirrored to support symmetric conversations. Secondly, our current display allows observers to see perspective-correct stereo images from multiple viewpoints. It could also support free viewpoints by tracking observers' positions. Thirdly, we hope to leverage our system for 3-way or N-way teleconferencing scenarios. Support of a teleconference with N users requires N $\times$ (N-1) data streams. Since avatar mediated interaction does not require significant bandwidth for transmission, our design would easily allow for such scaling. Lastly, an interesting question is the potential support for live multiple video streaming. We plan to further investigate on this topic, perhaps using a light field camera to capture the remote person or 360° array of cameras around the remote person.

We empirically evaluated the effect of perspective on the user's accuracy in judging gaze direction. The results revealed that parallax provides a dominant effect in improving the effectiveness with which users were able to estimate the gaze direction, with additional effect for motion parallax augmented by stereoscopy. The results also showed magnitude of the bias due to the lack of motion parallax and stereoscopic cues is less sensitive to vertical direction than horizontal direction.

8.4 Spherical avatar telepresence system

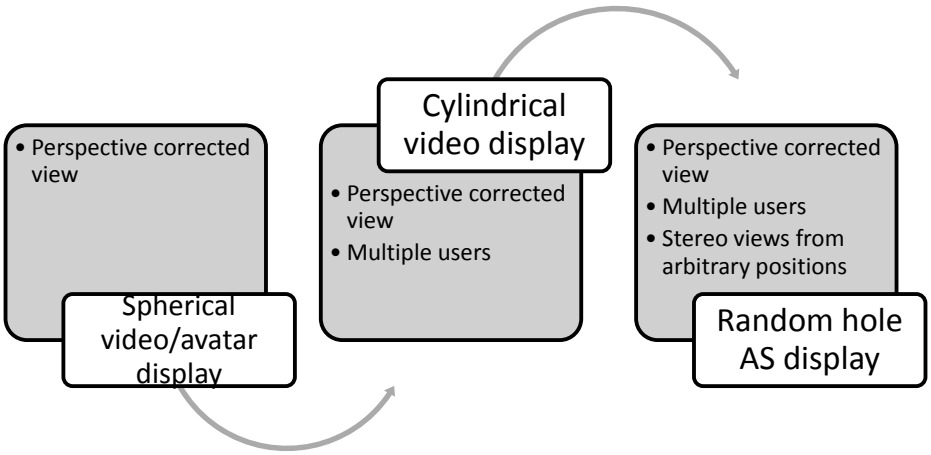
For the last system, we have presented a spherical display featuring a view-dependent rendering method to represent virtual avatars. We detailed a method for enabling the displayed avatar to reproduce the facial expression captured from a person in real-time and with high-fidelity. The system provided observers with perspective-correct rendering and the nature of the display offers surrounding visibility whereas flat displays are only viewable from the front. This borderless spherical display can be statically situated as an interesting display for virtual avatars or other content. It could also be mounted on a robot as a mobile display for telepresence.

We investigated the display in the context of a trust scenario. We investigated the effects of display type (sphere and flat) and viewing angle for trust assessments in avatar-mediated interaction. While participants were able to discriminate trustworthy and less trustworthy advisers irrespective of display type, a negative bias for flat display can interfere with users' ability to discriminate effectively. The interference became significant at off-center viewing angles, where the flat display no longer allows an undistorted and clear view. This demonstrates that a participant's level of trust can be manipulated during avatar-mediated communication by the appearance of a remote interactant.

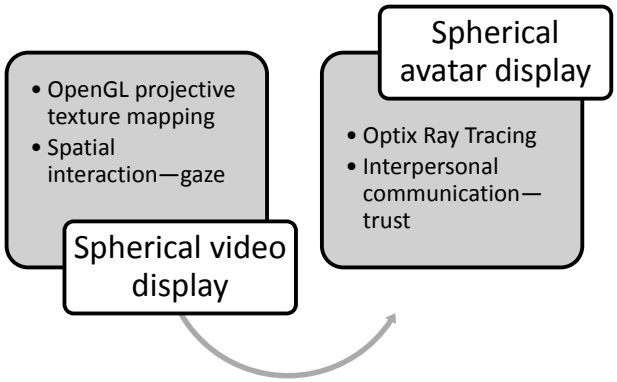
The surrounding characteristics of spherical displays allow perspective-correct imagery to be seen from all viewing directions, and hence avoid the problems that we have observed with traditional flat displays. By preserving a virtual avatar's correct appearance and gaze direction, the spherical display is able to maintain a consistently high level of trust regardless of viewing position.

8.5 Relationship among four different systems

In this thesis, we presented four telepresence systems, each of them has made a further contribution to improve teleconferencing experience. As presented in Figure 8.1(a), both the spherical video telepresence system and spherical avatar telepresence system provide 360 degree perspective-correct imagery for a single user. The cylindrical video telepresence system extend this function to multiple users. Furthermore, the random hole multiview telepresence system not only provide perspective-correct imagery for multiple users, but also support stereo views from arbitrary locations. Note that our



(a) Affordance of displays



(b) Rending & Evaluation scenarios

Figure 8.1: Relationship among four different telepresence systems.

current prototype of the random hole multiview telepresence system is based on a traditional flat display surface. Systems using similar principles could be configured to support a 360 degree view, by using a cylindrical display surface.

Figure 8.1(b) shows the relationship between the spherical video telepresence system and the spherical avatar telepresence system. Both of them used the same hardware, but the rending methods are different. For the spherical video telepresence system, we used the OpenGL polygonal rendering approach. However, we used a ray tracing engine that should provide higher quality images with less distortion. Additionally, we evaluated object focused gaze direction of the spherical display in the first experiment, but we investigated the surrounding features of spherical displays by using more complicated scenario: interpersonal trust.

8.6 Contribution

Video conferencing attempts to convey subtle cues of face-to-face interaction, but it is generally believed to be less effective than face to face. We argue that careful design based on an understanding of non-verbal communication can mitigate these differences. The overarching goal of the research was to improve teleconferencing experience to enable people to effectively accomplish the task at hand. Chapter 1 introduced five research motivations that are components of this goal.

For the first research motivation, many teleconferencing systems have been developed to support gaze awareness. However, the majority use a 2D planar display which is visible from the front only. In this research, we continue to push the boundaries on teleconferencing design. We introduced three novel situated displays, including the spherical video telepresence system (Section 3.1), the spherical avatar telepresence system (Section 3.2) and the cylindrical video telepresence system (Section 3.3), which offers a 360 view.

For the second and third research motivations, current immersive systems and situated displays can replicate a correct gaze direction. In most of these systems, the motion parallax is achieved by providing a perspective correct image via a single user's head position tracking. Eventually only one image is presented on the display. Thus they are currently developed for a single observer; other users can view the display but will see highly a distorted view. However, our cylindrical video telepresence system (Section 3.3) and random hole autostereoscopic multiview telepresence system (see Section 3.4) display present multiple images simultaneously and thus multiple users can see the correct view.

For the fourth research motivation, the use of autostereoscopic display technologies could support multiple users simultaneously each with their own perspective-correct view without the need for special eyewear. However, these are usually restricted to specific optimal viewing zones. In this thesis, our random hole autostereoscopic multiview telepresence system (see Section 3.4) provides perspective-correct stereoscopic imagery for multiple users in arbitrary positions.

For the fifth research motivation, gaze and trust formation on these situated displays has not been evaluated yet. In this thesis, three evaluations on the affordance of object focused gaze of telepresence displays together with one evaluation on the

affordance of interpersonal trust of telepresence display are documented throughout Chapter 4 to Chapter 7.

As the amount of our time spent in mediated interaction increases, these systems and findings of user studies discussed above have significant implications for teleconferencing in general.

For the contributions to telepresence displays, we have presented four new design of teleconferencing system that helps harness the power of nonverbal communication. In particular, our systems avoid the distortion of the gaze cues we have observed with traditional displays.

For the contributions to human factors, we introduced several empirical methods and performed studies based on these methods to improve our understanding of the technological and the social implications of our telepresence displays design. We began by evaluating the affordance of object focused gaze of telepresence displays and demonstrated the gaze-preserving capability of our displays. We then modelled an experimental study on a user-advisor relationship. Using this method, we investigated the influence of display type and viewing angle on how people place their trust during avatar-mediated interaction. The results showed how trust can be altered depending on how one views the avatar. This would contribute to a theoretical understanding of human, nonverbal communication and inform future design of communication technology.

For the contributions to graphical rendering, we developed view-dependent ray traced rendering methods for the spherical display and the random hole display. This could be extended to other display surfaces.

8.7 Directions for future work

For future work, besides the future directions of designs and evaluations of our four displays, discussed above (see Section 8.1 to Section 8.4), there are several routes for future research.

8.7.1 Future display technologies

8.7.1.1 Full-body telepresence display

This thesis has aimed to focus on the display of a head because when we interact with others we pay the most attention to the face. Faces are interesting because they convey eye gaze, expressions and gestures and are used as a central channel of communication. However, prior work, such as Ultra-Videoconferencing [26], suggests that to avoid misperceptions of social distance and to aid in a sense of realism, preservation of body size is important [74, 54]. Therefore, such investigation of producing life size is a potentially revealing avenue of research. Our situated displays could easily be integrated into a robotic platform to have a body. Also, we could to leverage our systems to produce life-size images by using larger higher resolution screens.

8.7.1.2 Telepresence robot, mobility and haptic feedbacks

Our situated displays could be mounted on a robot to include haptic (hands or body) or mobility capabilities. Current telepresence robots generally use flat screens, with a web-cam view of the remote participant. This web-cam view could be rendered on to a situated display and oriented, independent of the robot base, to face in any direction. This would support more rapid head movement than turning the base itself. This could help in social situations where attention needs to be directed quickly. The direction of this surface video view could be driven in multiple ways (e.g., similar to Animatronic Shader Lamps Avatars [61]).

8.7.1.3 Shape changing interfaces

To further enhance the teleconferencing experience, we are interested in exploring shape-changing displays. The main functional purpose of applying shape change is to better communicate information depending on the number and position of the observers. For example, we could adjust the cylindrical screen of the cylindrical video telepresence system to different shapes and sizes. we could also projecting live video onto a surface shaped like a human face. Also, dynamic affordances are another potentially useful feature, where shape change is used to communicate possibilities for action. Another functional purpose of shape change is to use it for providing haptic feedback. The haptic feedback could be used to create social presence by recording the interactions of one user and play them back either locally or on remotely placed de-

vice. We would like to build novel display surfaces that could be change dynamically or retain arbitrary shapes, and investigate how users may perceive them and different experiences that may engender.

8.7.2 Future user experience evaluations

8.7.2.1 Evaluating two-way or N-way conversation scenario

In this work, our current systems are focused on supporting one-way conversation scenarios. Section 8.1 and section 8.3 detailed how our systems could be extended to support two-way or N-way conversation scenario in video-mediated communication and avatar-mediated communication respectively. Once two way conversation had been established, the most obvious test would be evaluating the extent to which eye contact could be achieved. Additionally, as reviewed in chapter 2, many other potential studies could be used to assess whether the system could improve the sense of telepresence and effective communication.

8.7.2.2 Natural interaction scenario

This thesis introduced several frameworks for evaluating teleconferencing systems, which could be useful for the future system evaluation. Gaze, attention, eye contact, and trust are fundamental parts of human interaction, and we intend to explore other important scenarios and natural interaction in future work. We are also interested in developing evaluation methods range from assessing subjective phenomena (e.g., through questionnaires) to observing objective phenomena (e.g., by measuring biosignals).

Bibliography

[1] ACKER, S. R., AND LEVITT, S. R. Designing videoconference facilities for improved eye contact. *Journal of Broadcasting & Electronic Media 31*, 2 (1987), 181–191.

[2] ADALGEIRSSON, S. O., AND BREAZEAL, C. Mebot: a robotic platform for socially embodied presence. In *International Conference on Human-Robot Interaction* (2010), IEEE, pp. 15–22.

[3] AL MOUBAYED, S., EDLUND, J., AND BESKOW, J. Taming mona lisa: communicating gaze faithfully in 2d and 3d facial projections. *ACM Transactions on Interactive Intelligent Systems 1*, 2 (2012), 25.

[4] ANSTIS, S. M., MAYHEW, J. W., AND MORLEY, T. The perception of where a face or television'portrait'is looking. *The American journal of psychology 82*, 4 (1969), 474–489.

[5] AOKI, T., WIDOYO, K., SAKAMOTO, N., SUZUKI, K., SABURI, T., AND YASUDA, H. Monjunochie system: videoconference system with eye contact for decision making. *IEIC Technical Report (Institute of Electronics, Information and Communication Engineers) 98*, 552 (1999), 9–14.

[6] ARAI, H., KURIKI, M., AND SAKAI, S. New eye-contact technique for videophone. In *SID Symposium Digest of Technical Papers* (1992), SID, pp. 149–152.

[7] ARAI, J., OKUI, M., YAMASHITA, T., AND OKANO, F. Integral three-dimensional television using a 2000-scanning-line video system. *Applied optics 45*, 8 (2006), 1704–1712.

[8] ATZPADIN, N., KAUFF, P., AND SCHREER, O. Stereo analysis by hybrid recursive matching for real-time immersive video conferencing. *IEEE Transactions on Circuits and Systems for Video Technology 14*, 3 (2004), 321–334.

[9] AZUMA, R., BAILLOT, Y., BEHRINGER, R., FEINER, S., JULIER, S., AND MACINTYRE, B. Recent advances in augmented reality. *IEEE Computer Graphics and Applications 21*, 6 (2001), 34–47.

[10] BACHARACH, M., AND GAMBETTA, D. Trust as type detection. In *Trust and deception in virtual societies* (2001), Kluwer Academic Publishers, pp. 1–26.

[11] BAILENSON, J. N., YEE, N., MERGET, D., AND SCHROEDER, R. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments 15*, 4 (2006), 359–372.

[12] BARNETT WHITE, T. Consumer trust and advice acceptance: The moderating roles of benevolence, expertise, and negative emotions. *JCP 15*, 2 (2005), 141–148.

[13] BEKKERING, E., AND SHIM, J. Trust in videoconferencing. *Communications of the ACM 49*, 7 (2006), 103–107.

[14] BENKO, H., JOTA, R., AND WILSON, A. Miragetable: freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012), ACM, pp. 199–208.

[15] BILLINGHURST, M., POUPYREV, I., KATO, H., AND MAY, R. Mixing realities in shared space: An augmented reality interface for collaborative computing. In *IEEE International Conference on Multimedia and Expo* (2000), vol. 3, IEEE, pp. 1641–1644.

[16] BLINN, J., AND NEWELL, M. Texture and reflection in computer generated images. *Communications of the ACM 19* (Oct. 1976), 542–547.

[17] BLUNDELL, B. G., AND SCHWARZ, A. J. Volumetric three-dimensional display systems. *Volumetric Three-Dimensional Display Systems 1* (2000).

[18] BÖCKER, M., BLOHM, W., AND MÜHLBACH, L. Anthropometric data on horizontal head movements in videocommunications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (1996), ACM.

[19] BÖCKER, M., RUNDE, D., AND MÜHLBACH, L. On the reproduction of motion parallax in videocommunications. In *HFES* (1995), vol. 39, SAGE Publications.

[20] BOHANNON, L. S., HERBERT, A. M., PELZ, J. B., AND RANTANEN, E. M. Eye contact and video-mediated communication: A review. *Displays* (2012).

[21] BOLTON, J., KIM, K., AND VERTEGAAL, R. A comparison of competitive and cooperative task performance using spherical and flat displays. In *ACM conference on Computer supported cooperative work* (2012), ACM, pp. 529–538.

[22] BOS, N., OLSON, J., GERGLE, D., OLSON, G., AND WRIGHT, Z. Effects of four computer-mediated communications channels on trust development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2002), ACM, pp. 135–140.

[23] BUCHNER, G. Interactive audio-visual system, May 23 2006. US Patent 7,048,386.

[24] CHEN, W., BOUGUET, J., CHU, M., AND GRZESZCZUK, R. Light field mapping: efficient representation and hardware rendering of surface light fields. *ACM Transactions on Graphics 21*, 3 (2002), 447–456.

[25] COOK, M. Gaze and mutual gaze in social encounters: How longand whenwe look others" in the eye" is one of the main signals in nonverbal communication. *American Scientist* (1977), 328–333.

[26] COOPERSTOCK, J. R., ROSTON, J., AND WOSZCZYK, W. Broadband networked audio: Entering the era of multisensory data distribution. In *18th Internation Congress of Acoustics* (2004).

[27] CORNES, R., AND SANDLER, T. *The theory of externalities, public goods, and club goods*. Cambridge University Press, 1996.

[28] CRUZ-NEIRA, C., SANDIN, D., AND DEFANTI, T. Surround-screen projection-based virtual reality: the design and implementation of the cave. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques* (1993), ACM, pp. 135–142.

[29] DAVIS, J., FARNHAM, S., AND JENSEN, C. Decreasing online'bad'behavior. In *SIGCHI extended abstracts on Human factors in computing systems* (2002), ACM, pp. 718–719.

[30] DAWES, R. M. Social dilemmas. *Annual review of psychology 31*, 1 (1980), 169–193.

[31] DEUTSCH, M. Trust and suspicion. *The Journal of conflict resolution 2*, 4 (1958), 265–279.

[32] DUCHENEAUT, N., YEE, N., NICKELL, E., AND MOORE, R. J. Alone together?: exploring the social dynamics of massively multiplayer online games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2006), ACM, pp. 407–416.

[33] DUNBAR, D., AND HUMPHREYS, G. A spatial data structure for fast poisson-disk sample generation. In *ACM Transactions on Graphics* (2006), vol. 25, ACM, pp. 503–508.

[34] FLOOD, M. Some experimental games. *Management Science* (1958), 5–26.

[35] GEMMELL, J., TOYAMA, K., ZITNICK, C., KANG, T., AND SEITZ, S. Gaze awareness for video-conferencing: A software approach. *Multimedia 7*, 4 (2000), 26–35.

[36] GIBBS, S., ARAPIS, C., AND BREITENEDER, C. Teleport–towards immersive copresence. *Multimedia Systems 7*, 3 (1999), 214–221.

[37] GIBSON, J., AND PICK, A. Perception of another person's looking behavior. *The American Journal of Psychology 76*, 3 (1963), 386–394.

[38] GLASSNER, A. *An introduction to ray tracing*. Morgan Kaufmann, 1989.

[39] GREENE, N. Environment mapping and other applications of world projections. *IEEE Computer Graphics and Applications 6* (Nov. 1986), 21 –29.

[40] GROSS, M., WÜRMLIN, S., NAEF, M., LAMBORAY, E., SPAGNO, C., KUNZ, A., KOLLER-MEIER, E., SVOBODA, T., VAN G., L., LANG, S., STREHLKE, K., MOERE, A., AND STAADT, O. blue-c: a spatially immersive display and 3d video portal for telepresence. In *SIGGRAPH* (2003), ACM.

[41] HINDMARSH, J., FRASER, M., HEATH, C., BENFORD, S., AND GREENHALGH, C. Object-focused interaction in collaborative virtual environments. *ACM Transactions on Computer-Human Interaction 7*, 4 (2000), 477–509.

[42] HINRICHS, U., CARPENDALE, S., AND SCOTT, S. Evaluating the effects of fluid interface components on tabletop collaboration. In *Proceedings of the working conference on Advanced visual interfaces* (2006), ACM, pp. 27–34.

[43] HIRAYAMA, Y. One-dimensional integral imaging 3d display systems. In *Proceedings of the 3rd International Universal Communication Symposium* (2009), ACM, pp. 141–145.

[44] HORN, D. B., KARASIK, L., AND OLSEN, J. S. The effects of spatial and temporal video distortion on lie detection performance. In *SIGCHI Extended Abstracts on Human Factors in Computing Systems* (2002), ACM, pp. 714–715.

[45] JOAQUIM, P., AND DE SÁ, M. *Applied statistics using SPSS, STATISTICA and MATLAB*. Springer, 2003.

[46] JONES, A., LANG, M., FYFFE, G., YU, X., BUSCH, J., MCDOWALL, I., BOLAS, M., AND DEBEVEC, P. Achieving eye contact in a one-to-many 3d

video teleconferencing system. *ACM Transactions on Graphics 28*, 3 (2009), 64.

[47] JOUPPI, N. P., IYER, S., THOMAS, S., AND SLAYDEN, A. Bireality: mutually-immersive telepresence. In *Multimedia* (2004), ACM, pp. 860–867.

[48] KARNIK, A., MAYOL-CUEVAS, W., AND SUBRAMANIAN, S. Mustard: a multi user see through ar display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012), ACM, pp. 2541–2550.

[49] KHALILBEIGI, M., LISSERMANN, R., MÜHLHÄUSER, M., AND STEIMLE, J. Xpaaand: interaction techniques for rollable displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), ACM, pp. 2729–2732.

[50] KIM, K., BOLTON, J., GIROUARD, A., COOPERSTOCK, J., AND VERTEGAAL, R. Telehuman: effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012), ACM, pp. 2531–2540.

[51] KIM, S., CAO, X., ZHANG, H., AND TAN, D. Enabling concurrent dual views on common lcd screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012), ACM, pp. 2175–2184.

[52] KITAMURA, Y., KONISHI, T., YAMAMOTO, S., AND KISHINO, F. Interactive stereoscopic display for three or more users. In *SIGGRAPH* (2001), ACM, pp. 231–240.

[53] KLEINKE, C. L. Gaze and eye contact: a research review. *Psychological bulletin 100*, 1 (1986), 78.

[54] KLEINSMITH, A., AND BIANCHI-BERTHOUZE, N. Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing 4*, 1 (2013), 15–33.

[55] KOLLOCK, P. Social dilemmas: The anatomy of cooperation. *Annual review of sociology* (1998), 183–214.

[56] KOOIMA, R., PRUDHOMME, A., SCHULZE, J., SANDIN, D., AND DEFANTI, T. A multi-viewer tiled autostereoscopic virtual reality display. In *proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology* (2010), ACM, pp. 171–174.

[57] KULIK, A., KUNERT, A., BECK, S., REICHEL, R., BLACH, R., ZINK, A., AND FROEHLICH, B. C1x6: a stereoscopic six-user display for co-located collaboration in shared virtual environments. In *ACM Transactions on Graphics* (2011), vol. 30, ACM, p. 188.

[58] LANTZ, E. Future directions in visual display systems. *SIGGRAPH 31*, 2 (1997), 38–42.

[59] LEE, M. K., AND TAKAYAMA, L. Now, i have a body: Uses and social norms for mobile remote presence in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), ACM, pp. 33–42.

[60] LI, H., YU, J., YE, Y., AND BREGLER, C. Realtime facial animation with on-the-fly correctives. *ACM Transactions on Graphics 32*, 4 (2013), 42.

[61] LINCOLN, P., WELCH, G., NASHEL, A., ILIE, A., FUCHS, H., ET AL. Animatronic shader lamps avatars. In *IEEE International Symposium on Mixed and Augmented Reality* (2009), IEEE, pp. 27–33.

[62] LUCENTE, M. Interactive three-dimensional holographic displays: seeing the future in depth. *SIGGRAPH 31*, 2 (1997), 63–67.

[63] MATSUYAMA, T., AND TAKAI, T. Generation, visualization, and editing of 3d video. In *International Symposium on 3D Data Processing Visualization and Transmission* (2002), IEEE, pp. 234–245.

[64] MATUSIK, W., AND PFISTER, H. 3d tv: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. In *ACM Transactions on Graphics* (2004), vol. 23, ACM, pp. 814–824.

[65] MAYER, R. C., DAVIS, J. H., AND SCHOORMAN, F. D. An integrative model of organizational trust. *Academy of management review* (1995), 709–734.

[66] MULLIGAN, J., ISLER, V., AND DANIILIDIS, K. Trinocular stereo: A real-time algorithm and its evaluation. *International Journal of Computer Vision 47*, 1-3 (2002), 51–61.

[67] MURRAY, N., ROBERTS, D., STEED, A., SHARKEY, P., DICKERSON, P., AND RAE, J. An assessment of eye-gaze potential within immersive virtual environments. *ACM Transactions on Multimedia Computing, Communications, and Applications 3*, 4 (2007), 8.

[68] NAGANO, K., JONES, A., LIU, J., BUSCH, J., YU, X., BOLAS, M., AND DEBEVEC, P. An autostereoscopic projector array optimized for 3d facial display. In *SIGGRAPH* (2013), ACM, p. 3.

[69] NASHEL, A., AND FUCHS, H. Random hole display: A non-uniform barrier autostereoscopic display. In *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2009* (2009), IEEE, pp. 1–4.

[70] NEGROPONTE, N. being digital. alfred a, 1995.

[71] NGUYEN, D., AND CANNY, J. Multiview: spatially faithful group video conferencing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA, 2005), ACM, pp. 799–808.

[72] NGUYEN, D. T., AND CANNY, J. Multiview: improving trust in group video conferencing through spatial faithfulness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2007), ACM, pp. 1465–1474.

[73] NORRIS, J., SCHNÄDELBACH, H., AND QIU, G. Camblend: an object focused collaboration tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012), ACM, pp. 627–636.

[74] NOWAK, K., AND BIOCCA, F. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments 12*, 5 (2003), 481–494.

[75] OKADA, K., MAEDA, F., ICHIKAWAA, Y., AND MATSUSHITA, Y. Multiparty videoconferencing at virtual social distance: Majic design. In *ACM conference on Computer supported cooperative work* (Chapel Hill, North Carolina, USA, 1994), ACM, pp. 385–393.

[76] OKADA, K.-I., MAEDA, F., ICHIKAWAA, Y., AND MATSUSHITA, Y. Multiparty videoconferencing at virtual social distance: Majic design. In *ACM conference on Computer supported cooperative work* (1994), ACM, pp. 385–393.

[77] OLSON, G. M., AND OLSON, J. S. Distance matters. *Human-computer interaction 15*, 2 (2000), 139–178.

[78] OYEKOYA, O., STEPTOE, W., AND STEED, A. Sphereavatar: a situated display to represent a remote collaborator. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012), ACM, pp. 2551–2560.

[79] PAN, Y., OYEKOYA, O., AND STEED, A. A surround video capture and presentation system for preservation of eye-gaze in teleconferencing applications. *PRESENCE: Teleoperators and Virtual Environments 24*, 1 (2015), 24–43.

[80] PAN, Y., AND STEED, A. Preserving gaze direction in teleconferencing using a camera array and a spherical display. In *IEEE 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video* (2012), IEEE, pp. 1–4.

[81] PAN, Y., AND STEED, A. A gaze-preserving situated multiview telepresence system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2014), ACM, pp. 2173–2176.

[82] PAN, Y., STEPTOE, W., AND STEED, A. Comparing flat and spherical displays in a trust scenario in avatar-mediated interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2014), ACM, pp. 1397–1406.

[83] PARKER, S. G., BIGLER, J., DIETRICH, A., FRIEDRICH, H., HOBEROCK, J., LUEBKE, D., MCALLISTER, D., MCGUIRE, M., MORLEY, K., ROBISON, A., ET AL. Optix: A general purpose ray tracing engine. *ACM Transactions on Graphics 29*, 4 (2010), 66.

[84] PAULOS, E., AND CANNY, J. Prop: personal roving presence. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (1998), ACM, pp. 296–303.

[85] PERLIN, K., PAXIA, S., AND KOLLIN, J. S. An autostereoscopic display. In *SIGGRAPH* (2000), ACM, pp. 319–326.

[86] PETERKA, T., KOOIMA, R. L., SANDIN, D. J., JOHNSON, A., LEIGH, J., AND DEFANTI, T. A. Advances in the dynallax solid-state dynamic parallax barrier autostereoscopic visualization display system. *IEEE Transactions on Visualization and Computer Graphics 14*, 3 (2008), 487–499.

[87] PRINCE, S., CHEOK, A. D., FARBIZ, F., WILLIAMSON, T., JOHNSON, N., BILLINGHURST, M., AND KATO, H. 3d live: Real time captured content for mixed reality. In *IEEE International Symposium on Mixed and Augmented Reality* (2002), IEEE, pp. 7–317.

[88] RADOVAN, M., AND PRETORIUS, L. Facial animation in a nutshell: past, present and future. In *Proceedings of the 2006 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries* (2006), South African Institute for Computer Scientists and Information Technologists, pp. 71–79.

[89] RAE, I., TAKAYAMA, L., AND MUTLU, B. In-body experiences: embodiment, control, and trust in robot-mediated communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), ACM, pp. 1921–1930.

[90] RASKAR, R., WELCH, G., CUTTS, M., LAKE, A., STESIN, L., AND FUCHS, H. The office of the future: a unified approach to image-based modeling and spatially immersive displays. In *SIGGRAPH* (1998), ACM.

[91] RIEGELSBERGER, J., SASSE, M. A., AND MCCARTHY, J. D. The mechanics of trust: a framework for research and design. *International Journal of Human-Computer Studies 62*, 3 (2005), 381–422.

[92] RIEGELSBERGER, J., SASSE, M. A., AND MCCARTHY, J. D. Rich media, poor judgement? a study of media effects on users trust in expertise. In *People and Computers XIXThe Bigger Picture*. Springer, 2006, pp. 267–284.

[93] ROBERTS, D., WOLFF, R., OTTO, O., AND STEED, A. Constructing a gazebo: supporting teamwork in a tightly coupled, distributed task in virtual reality. *Presence: Teleoperators and Virtual Environments 12*, 6 (2003), 644–657.

[94] ROBERTS, D., WOLFF, R., RAE, J., STEED, A., ASPIN, R., MCINTYRE, M., PENA, A., OYEKOYA, O., AND STEPTOE, W. Communicating eye-gaze across a distance: Comparing an eye-gaze enabled immersive collaborative virtual environment, aligned video conferencing, and being together. In *IEEE Virtual Reality* (Washington, DC, USA, 2009), IEEE, pp. 135–142.

[95] ROBERTS, D. J., RAE, J., DUCKWORTH, T. W., MOORE, C. M., AND ASPIN, R. Estimating the gaze of a virtuality human. *IEEE Transactions on Visualization and Computer Graphics 19*, 4 (2013), 681–690.

[96] ROBERTS, D. J., WOLFF, R., AND OTTO, O. Supporting a closely coupled task between a distributed team: using immersive virtual reality technology. *Computing and Informatics 24*, 1 (2012), 7–29.

[97] ROUDAUT, A., KARNIK, A., LÖCHTEFELD, M., AND SUBRAMANIAN, S. Morphees: toward high shape resolution in self-actuated flexible mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), ACM, pp. 593–602.

[98] SADAGIC, A., TOWLES, H., LANIER, J., FUCHS, H., VAN DAM, A., DANILIDIS, K., MULLIGAN, J., HOLDEN, L., AND ZELEZNIK, B. National tele-immersion initiative: Towards compelling tele-immersive collaborative environments. In *Presentation given at Medicine meets Virtual Reality conference* (2001).

[99] SAITO, H., BABA, S., KIMURA, M., VEDULA, S., AND KANADE, T. Appearance-based virtual view generation of temporally-varying events from

multi-camera images in the 3d room. In *International Conference on 3-D Digital Imaging and Modeling* (1999), IEEE, pp. 516–525.

[100] SAKAMOTO, D., KANDA, T., ONO, T., ISHIGURO, H., AND HAGITA, N. Android as a telecommunication medium with a human-like presence. In *International Conference on Human-Robot Interaction* (2007), IEEE, pp. 193–200.

[101] SANDIN, D. J., MARGOLIS, T., GE, J., GIRADO, J., PETERKA, T., AND DEFANTI, T. A. The varrier tm autostereoscopic virtual reality display. In *ACM Transactions on Graphics* (2005), vol. 24, ACM, pp. 894–903.

[102] SCHMEIL, A., AND BROLL, W. Mara-a mobile augmented reality-based virtual assistant. In *IEEE Virtual Reality* (2007), IEEE, pp. 267–270.

[103] SCHMEIL, A., EPPLER, M. J., AND DE FREITAS, S. A structured approach for designing collaboration experiences for virtual worlds. *Journal of the Association for Information Systems 13*, 10 (2012), 836–860.

[104] SCHREER, O., CHANG, K., HENDRIKS, E., SCHRAAGEN, J., STONE, J., TRUCCO, E., AND JEWELL, M. Virtual team user environments–a key application in telecommunication. In *Proc. of eBusiness and eWork* (2002), Citeseer.

[105] SCHREER, O., KAUFF, P., AND SIKORA, T. *3D videocommunication*. Wiley Online Library, 2005.

[106] SCHWESIG, C., POUPYREV, I., AND MORI, E. Gummi: a bendable computer. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2004), ACM, pp. 263–270.

[107] SEGAL, M., KOROBKIN, C., VAN WIDENFELT, R., FORAN, J., AND HAEBERLI, P. Fast shadows and lighting effects using texture mapping. In *SIGGRAPH* (1992), vol. 26, ACM, pp. 249–252.

[108] SEKITOH, M. Bird's eye view system for its. In *IEEE Intelligent Vehicle Symposium* (2001), IEEE.

[109] SELLEN, A., BUXTON, B., AND ARNOTT, J. Using spatial cues to improve videoconferencing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (1992), ACM, pp. 651–652.

[110] SELLEN, A., BUXTON, B., AND ARNOTT, J. Using spatial cues to improve videoconferencing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA, 1992), ACM, pp. 651–652.

[111] SLATER, M., STEED, A., AND CHRYSANTHOU, Y. *Computer graphics and virtual environments: from realism to real-time*. Addison Wesley, 2002.

[112] SLOVÁK, P., TROUBIL, P., AND HOLUB, P. Gcoll group-to-group videoconferencing system: design and first experiences. In *International Conference on Collaborative Computing: Networking, Applications and Worksharing* (2009), IEEE, pp. 1–9.

[113] SMOLIC, A., MÜLLER, K., DROESE, M., VOIGT, P., AND WIEGAND, T. Multiple view video streaming and 3d scene reconstruction for traffic surveillance. In *European Workshop on Image Analysis for Multimedia Interactive Services* (2003), pp. 427–432.

[114] STAVNESS, I., LAM, B., AND FELS, S. pcubee: a perspective-corrected handheld cubic display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2010), ACM, pp. 1381–1390.

[115] STEPTOE, W., OYEKOYA, O., MURGIA, A., WOLFF, R., RAE, J., GUIMARAES, E., ROBERTS, D., AND STEED, A. Eye tracking for avatar eye gaze control during object-focused multiparty interaction in immersive collaborative virtual environments. In *IEEE Virtual Reality* (2009), IEEE, pp. 83–90.

[116] STEPTOE, W., STEED, A., ROVIRA, A., AND RAE, J. Lie tracking: social presence, truth and deception in avatar-mediated telecommunication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2010), ACM, pp. 1039–1048.

[117] SWOL, L. M., AND SNIEZEK, J. A. Factors affecting the acceptance of expert advice. *British Journal of Social Psychology 44*, 3 (2005), 443–461.

[118] TANG, A., PAHUD, M., INKPEN, K., BENKO, H., TANG, J., AND BUXTON, B. Three's company: understanding communication channels in three-way distributed collaboration. In *ACM conference on Computer supported cooperative work* (2010), ACM, pp. 271–280.

[119] TANGER, R., KAUFF, P., AND SCHREER, O. Immersive meeting point. In *PCM* (2004), Springer-Verlag.

[120] TANIKAWA, T., SUZUKI, Y., HIROTA, K., AND HIROSE, M. Real world video avatar: real-time and real-size transmission and presentation of human figure. In *Proceedings of the 2005 international conference on Augmented tele-existence* (2005), ACM, pp. 112–118.

[121] TANIMOTO, M. Overview of free viewpoint television. *Signal Processing: Image Communication 21*, 6 (2006), 454–461.

[122] TEN KOPPEL, M., BAILLY, G., MÜLLER, J., AND WALTER, R. Chained displays: Configurations of public displays can be used to influence actor-, audience-, and passer-by behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012), ACM.

[123] TEOH, C., REGENBRECHT, H., AND O'HARE, D. Investigating factors influencing trust in video-mediated communication. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction* (2010), ACM, pp. 312–319.

[124] TROJE, N., AND SIEBECK, U. Illumination-induced apparent shift in orientation of human heads. *PERCEPTION-LONDON- 27* (1998), 671–680.

[125] TSUI, K., DESAI, M., YANCO, H., AND UHLIK, C. Exploring use cases for telepresence robots. In *International Conference on Human-Robot Interaction* (2011), ACM, pp. 11–18.

[126] VERTEGAAL, R., SLAGTER, R., VAN DER VEER, G., AND NIJHOLT, A. Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2001), ACM, pp. 301–308.

[127] VERTEGAAL, R., WEEVERS, I., SOHN, C., AND CHEUNG, C. Gaze-2: conveying eye contact in group video conferencing using eye-controlled camera direction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2003), ACM, pp. 521–528.

[128] VERTEGAAL, R., WEEVERS, I., SOHN, C., AND CHEUNG, C. Gaze-2: conveying eye contact in group video conferencing using eye-controlled camera direction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA, 2003), ACM, pp. 521–528.

[129] WEISE, T., BOUAZIZ, S., LI, H., AND PAULY, M. Realtime performance-based facial animation. *ACM Transactions on Graphics 30*, 77 (2011), 1–10.

[130] WILBURN, B., SMULSKI, M., LEE, K., AND HOROWITZ, M. The light field video camera. Tech. rep., DTIC Document, 2000.

[131] WILLIAMS, E. Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin 84*, 5 (1977), 963.

[132] WILLIAMS, L. Performance-driven facial animation. In *SIGGRAPH* (1990), vol. 24, ACM, pp. 235–242.

[133] WILSON, H., WILKINSON, F., LIN, L., AND CASTILLO, M. Perception of head orientation. *Vision research 40*, 5 (2000), 459–472.

[134] YE, G., FUCHS, H., ET AL. A practical multi-viewer tabletop autostereoscopic display. In *IEEE International Symposium on Mixed and Augmented Reality* (2010), IEEE, pp. 147–156.

[135] YENDO, T., FUJII, T., TANIMOTO, M., AND PANAHPOUR TEHRANI, M. The seelinder: Cylindrical 3d display viewable from 360 degrees. *Journal of visual communication and image representation 21*, 5-6 (2010), 586–594.

[136] ZHANG, Z. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 11 (2000), 1330–1334.

[137] ZHENG, J., VEINOTT, E., BOS, N., OLSON, J., AND OLSON, G. Trust without touch: jumpstarting long-distance trust with initial social activities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2002), ACM, pp. 141–146.